THERAPISTS' EXPERIENCE OF DIFFICULTIES IN THEIR PRACTICE

Volume I of II

Thomas Albert Schröder

Thesis submitted for the degree of Doctor of Philosophy

University of Warwick

Department of Psychology

September 1998

TABLE OF CONTENTS:

VOLUME I

VOLUME II

APPENDICES

LIST OF TABLES:

ACKNOWLEDGEMENTS

Writing a thesis can seem a solitary pursuit. Many people were instrumental, however, in bringing about the final result, and I gratefully acknowledge their contributions.

The greatest debt of gratitude I owe to all those psychotherapist colleagues who not only generously gave of their time to fill in overly long questionnaires, but who were also prepared to share difficult and sometimes painful moments from their practice. Without their openness and co-operation this study would not have got beyond the drawing board.

I am grateful to the other members of the Warwick Collaborative Group - Marc Binns, John and Marcia Davis, Robert Elliott, Valerie Francis, and Jim Kelman - for the initial inspiration and investment in the topic of therapist difficulties and for their continuing interest, even when the data turned out to look like a map of Lincolnshire.

John, you have supported me throughout this project, in sickness and in health, with your friendship, your clear thinking, and your unwavering attention to detail. Thank you!

Thank you also, friends from the Collaborative Research Network, especially Hansruedi Ambühl, Jean-François Botermans, Manfred Cierpka, David Orlinsky, Helge Rønnestad, and Ulrike Willutzki, for sustaining my enthusiasm when local support was hard to come by. Meetings with you have regularly restored my resolve and demonstrated to me that engaging in serious research and enjoying good things in life are not mutually exclusive. Special thanks to you, Hansruedi, for persuading me that I had something worthwhile to say, and helping me to say it in the place where I grew up.

I am grateful to my employers, Southern Derbyshire Mental Health (NHS) Trust, for tolerating this research, even though it did not generate clinical activity, and for providing logistical support. Special thanks are due to Luise Weigel, who helped with the early stages, and to Sue Osmond, who tirelessly tracked down obscure references and imperilled her budget by ordering Inter Library Loans in bulk.

A major part of this thesis relied on reliable ratings. For these, I am grateful to Marc Binns, John Davis, Simon Elliot, John Fletcher, Val Francis, Joe Kiff, Liza Monaghan, Nigel Runcorn, and Simon Thomas. Thank you, Paul Gilbert, Geoff Heath, and Ian Morley for reading and commenting on the final draft.

And, finally, my warmest thanks go to my family - Valerie, Elizabeth and Hannah - for being constant reminders that, while there are many important things to do in mid-life, writing a PhD thesis is not necessarily one of them.

DECLARATION:

Some of the material contained in this thesis has been published during the period of registration. The re-analysis of previous survey data reported on p. 12 has been included in

Schröder, T. A. (1997). Was lange währt wird auch nicht leichter - Schwierigkeiten und Bewältigungsstrategien in der psychotherapeutischen Arbeit. In P. L. Janssen, M. Cierpka, & P. Buchheim (Eds.), *Psychotherapie als Beruf.* Goettingen: Vandenhoeck & Ruprecht.

which is included in volume II as appendix 15.

Work in progress has also been published in conference abstracts as follows:

Schröder, T. A. (1993). *Varieties of therapists' experiences of difficulties in practice.* Paper presented at the Society for Psychotherapy Research, 4th European Meeting, Budapest, Hungary.

Schröder, T. A. & Davis, J. D. (1993,). *The new and the short and the tall. Varieties of therapists' experiences of difficulties.* Paper presented at the Society for Psychotherapy Research (UK), 10th Annual Meeting, Ravenscar.

Schröder, T. A. & Davis, J. D. (1994). *Dimensions of difficulties. How can psychotherapists' experiences of problems in their practice be stratified?* Paper presented at the Society for Psychotherapy Research, 25th Annual Meeting, York, UK.

Schröder, T. A., Willutzki, U., & Ambuehl, H. R. (1994). *Schwierigkeiten und Bewältigungsstrategien in der psychotherapeutischen Arbeit.* Paper presented at the Deutsche Gesellschaft für Psychologie, 39. Jahreskongress, Hamburg.

Schröder, T. A. & Davis, J. D. (1997). C*onspicuous by their presence - universality and specificity of therapists' experience of difficulties in their practice.* Paper presented at the Society for Psychotherapy Research, 28th Annual Meeting, Geilo, Norway.

Schröder, T. A. & Davis, J. D. (1998). W*hat is difficult for which therapist and why?* Paper presented at the Society for Psychotherapy Research, 29th Annual Meeting, Snowbird, UT.

ABSTRACT

This thesis studies psychotherapists' experiences of difficulties in practice. Building on previous work establishing *content, prevalence,* and *relationship to professional development* of difficulties, it reports three studies based on two surveys investigating dimensions and correlates of the *pervasiveness* of such experiences, both within (*specificity*) and between therapists (*universality*). Difficulties were initially conceptualised on a spectrum ranging from 'transient' (based on skills deficits) to 'paradigmatic' (based on therapists' enduring idiosyncratic characteristics). In a first survey, 57 British psychotherapists each contributed two written accounts of difficulties and completed measures developed to tap emotional concomitants of these experiences. The hypothesised co-occurence of extreme levels of these measures was not observed. In a second study using the same data, written accounts were rated in terms of theoretically refined definitions of transient and paradigmatic difficulty types, joined by a third, 'situational' category, denoting difficulties arising from external sources and universally experienced as problematic. Separate studies established reliability and validity of ratings. A further study based on a survey of 100 British and 30 German speaking therapists confirmed reliability of the rating system and, using difficulty types as dependent variables, validated them through their predicted associations with measures of competency deficits, therapists' internal conflicts, and indices of therapist/patient similarity; using newly developed as well as established (IIP, SASB) measures. Situational difficulties were associated with difficult patients but did not dislodge therapists from their professional stance. Transient difficulties were related to self-reported competency deficits, occurred earlier in therapists' careers and were experienced as unfamiliar. Paradigmatic difficulties were found to be idiosyncratic, have high impact, evoke negative internal states in therapists, arouse their wish for hostile control in the absence of perceived patient hostility, and relate to therapists' internal attachment conflicts and indices of similarity with their patients. Consequences for supervision and models of therapist development are discussed.

1. Chapter One: Introduction

Conducting psychotherapy is inherently difficult: working with distressed people can be distressing, attending to disturbed people is often disturbing and the mental pain of others frequently reminds us of our own. Psychotherapists generally seek to master difficult emotions by trying to understand them; researchers habitually deal with troublesome phenomena by trying to measure them.

Being both a psychotherapist and a researcher, striving to make sense of the difficulties which therapists experience in their daily practice contains for me elements of not only identification but also adaptation. No doubt, the interest in this topic has arisen out of my own therapeutic difficulties as much as out of learning about the struggles of other therapists as their supervisor and colleague. Investigating the experiences of others thus acquires direct relevance for understanding my own; after all, as Hunt (1989) reminds us, research "is, in part, the discovery of the self through the detour of the other" (p.42). But beyond this interest, researching therapists' difficulties - especially when using quantitative methodology - is also an attempt at mastery, at replacing not knowing with the temporary reassurance of a mathematical model[1].

The studies underpinning this thesis have produced a great many numbers. They have also elicited a vast amount of words - altogether nearly 400 narratives of difficulty episodes. Although the textual data were eventually converted into numerical variables, by 'assigning numbers to objects' (Nunnally, 1978, p. 7), I believe that

[1] Mittwoch (1990) puts similar sentiments into words when she comments on a study attempting to compute all possible transference permutations in a psychotherapy group: "I actually enjoyed reading [the] article, not least because like the author I sometimes indulge myself with a little mathematics, especially when I am anxious or finding it difficult to tolerate uncertainties" (p. 326).

words and their meanings were pivotal in helping participating therapists to generate internal representations of their experiences. They were certainly crucial in helping me to develop my understanding of the data I was encountering.

Whether psychotherapeutic topics are better investigated by concentrating on words or numbers, in other words by qualitative or quantitative methodology, is the subject of much current controversy. While this thesis needed both, it seems to me that in general the polarised debate between the two positions ultimately only serves to obscure that both words and numbers are but poor approximations to experience.

Although the concepts developed in this thesis are not bound to any particular theoretical persuasion, it would be surprising (and somewhat alarming) if they had not been influenced by the psychodynamic/humanistic/interpersonal framework which generally informs my thinking and clinical practice. One corollary of such a bias is the conviction that personal and professional life are inextricably intertwined[2]. How both the conjunction and separation of personal and professional development can be understood through the study of therapists' difficulties is a point taken up in the conclusion in chapter six.

[2] The close attention psychotherapists pay to boundaries is testimony to this nexus.

1.1 Background

As a discrete topic, therapists' experiences of trials and tribulations in their day-to-day practice have not attracted much attention in the clinical literature and have met with similar indifference in the research literature. There are, of course, abundant descriptions of therapeutic troubles to be found in clinical case studies. These do, however, usually focus on particular patient predicaments, problematic presentations, or pathological processes, rather than on the qualities of their experiential concomitants as seen from the therapist's point of view. The therapist is more clearly in focus in the (almost exclusively psychodynamic) clinical literature on countertransference, which contains some very vivid descriptions of analytic asperities (for instance Searles, 1979). Yet, the concept of countertransference is much more inclusive than that of difficulties, encompassing unproblematic experiences as well as troublesome but unconscious reactions. Both clinical and empirical explorations of countertransference therefore touch tangentially on the topic of this thesis without speaking directly to its main concern - the conscious experience of difficulties in the practice of psychotherapy, regardless of their source.

There are some anecdotal collections of therapist difficulties, portraying psychotherapy as an 'impossible profession' (Malcolm, 1988) or at least 'a perilous calling' (Sussman, 1995), an idea that (in its Californian variant) has even found its way into popular fiction (Kellerman, 1985). Apart from these few exceptions, however, the discussions of therapists' experiences of difficulty appear confined to the informal setting of the coffee break, where those who are fortunate enough to work within a supportive team can find solace and enlist sympathy; or to the more

circumscribed context of supervision or intervision, where troubled therapists can at least expect to have a hearing and maybe to elicit constructive criticism.[3]

This relative public neglect of a topic which frequently preoccupies practitioners privately probably reflects an ambivalence in the clinical field as to how therapists' difficulties should be regarded. One extreme position would look at difficulties as the consequence of technical mistakes - betraying therapists' inexperience or lack of competence - which threaten to disrupt process and to vitiate outcome. The antithetical extreme position would view difficulties as integral to therapy, inevitably occurring and potentially informative through reproducing patients' intrapersonal and interpersonal difficulties in the consulting room. In the psychodynamic literature this controversy might be reflected by opinions as diverse as those of Kernberg (1984) and Langs (1973), who hold that mis-timing an interpretation inevitably leads to a disruption of the therapeutic process, accompanied by attenuated or at least delayed therapeutic effectiveness, and that of Coltart (1986) and Symington (1983), whose investment in 'not knowing' would lead them to assume that a mis-timed interpretation might well reflect the absence of a state of readiness in the patient to make use of such an intervention, and therefore would be diagnostic rather than disruptive. Within the systemic literature, there is an analogous debate as to whether it is necessary for therapists to be drawn into a family's interactional dynamics in order to 'join' them, or whether being dislodged from an outsider's position decreases their

---

[3] That this latter expectation cannot always be relied upon is vividly illustrated by Rioch, Coulter, & Weinberger (1976), who quote a member of one of their supervision groups commenting thus on the influence of that group on his experience of a difficulty : "I kept thinking that Ellen would say it had all been very superficial and Doug would say it was unclear and that I seemed to him very confused. I don't know what you would say, Dr. Rioch, except that I shouldn't have let her go over time. Gail would say I couldn't work with her if I didn't know what she wanted. And everybody would say she had walked over me, so I kept her on and practically shouted at her that I had to find out what her problem was. She got quite annoyed and said she had been telling me about it all this time. So I said "We have to stop now. Tell me next time so I can understand it." And I got up and walked out."

leverage on the system and diminishes their effectiveness (Palazzoli-Selvini, Boscolo, Cecchin, & Prata, 1980; Cecchin, 1987; Real, 1990).

Ambivalence about therapists' difficulties in the clinical field is probably matched by misgivings about the topic in the research community. While over the past forty years several phases of psychotherapy research may be distinguished (Barkham et al., 1998), none of them has paid much attention to the therapist, other than as an assortment of variables potentially affecting outcome (which have been extensively studied - see 2.2.2 below). The notion of a legitimate field of enquiry called 'psychotherapist research' is quite recent, and the available material is as yet sparse (see 2.2.1 below). As Orlinsky & Howard (1977) observed more than 20 years ago, it seems almost as if psychotherapy research is only regarded as justifiable if it focuses on process and outcome for clients, while being perceived as close to self-indulgence if it takes the therapist's perspective on process and outcome in its sights. If the latter has received attention at all, the link to potentially negative client outcomes has been prominent, such as in studies of therapist burnout (Maslach, 1982), therapist distress (Beutler, Crago, & Arizmendi, 1986) or therapist technical mistakes (Meares & Hobson, 1977).

As for the more circumscribed topic of therapists' subjective experience of difficulties in their practice, ambivalence is also noticeable in the reactions of colleagues who find themselves confronted with the concept. Having conducted several workshops on the topic, I have noticed amongst participants a dual reaction - on the one hand a certain coyness about admitting to the experience that conducting therapy is not always plain sailing, and on the other a palpable relief that this experience is not only shared but can also be validated by others and therefore rendered legitimate. This mixed reaction is also reflected in some of the comments made by respondents to the

two questionnaires which provided the data base for this thesis, and maybe it is not altogether surprising in a profession where practitioners may at least be partially motivated by the desire to resolve some of their own problems by treating them in others.

The relationship of difficulties to outcome does, in my view, ultimately matter. There are, however, two outcomes - one for the therapist and one for the patient - which need to be clarified and understood. Understanding the therapist's side of the equation requires both a grasp of the phenomenology of experiences of difficulty and a heuristic to generate theoretical formulations. Both of these are, at present, largely absent from the literature.

The lack of structured investigations into the area of therapist difficulties was one of the factors inspiring a collaborative group, meeting at the University of Warwick (and including the present author), to undertake such a study. Their work is reviewed in detail in 1.2 below. Other investigations, which have a more peripheral bearing on the topic, are reviewed in the next chapter.

1.2 The Work of the Warwick Collaborative Group

A study of therapists' descriptions of difficulties and associated coping efforts was conducted between 1984 and 1987 by a group of seven clinician researchers meeting at the University of Warwick. This inquiry took its point of departure from a qualitative self-study, developing descriptive categories of therapist difficulties and coping strategies using a 'grounded theory' approach (Glaser & Strauss, 1967), from material supplied by the investigators. After several iterative steps of revising categories after applying them to fresh descriptive material, two taxonomies resulted (Davis et al., 1987, 1989). The draft final version of the taxonomy of therapist difficulties comprised the following ten categories[4]:

1. Therapist Incompetent
2. Therapist Damaging
3. Therapist Perplexed (Technical Puzzlement) *(later called Therapist Puzzled)*
4. Therapist Threatened
5. Therapist Out Of Rapport
6. Therapist's Personal Issues
7. Painful Reality
8. Therapist Stuck
9. Therapist Thwarted
10. Moral/Ethical Dilemma *(previously merged with Painful Reality)*

The conceptual basis for each category is a phenomenological description of the therapist's experience rather than a behavioural description of the therapist's or the

[4] see appendix 1 for a manual with definitions and sample difficulty accounts (Binns et al., 1986).

patient's actions as the latter could be given different meanings by different therapists.

Subsequent studies demonstrated that the taxonomy could be used reliably (Cronbach's alphas (Cronbach, 1951) for seven raters ranging from .88 to .95) to categorise difficulty accounts generated by other psychotherapists and that it could also reliably be applied by raters who were not psychotherapists, but who were experienced in mental health work (in this case psychiatric nurses). Finally, a revised version of the taxonomy was used to rate 286 written accounts generated by 96 British psychotherapists who had responded to a survey questionnaire (Schröder et al., 1987b) which asked each respondent to supply three accounts of difficult situations under the headings

1. What did you or your patient (client) do which made the situation difficult?
2. What feelings or personal reactions did you experience in the situation?
3. How did you attempt to deal or cope with this difficulty?
4. How did this situation turn out?

while also eliciting information about a variety of therapist-, patient-, and situational variables. Judges rated the presence of each of the taxonomy categories in each of the written accounts on a 0 - 3 (definitely absent - definitely present) scale and nominated one of the categories, which best categorised the difficulty, as 'predominant' (see appendix 1). Reliabilities were again satisfactory with Cronbach's alphas (calculated for six raters) ranging from .85 for the category "Painful Reality" to .93 for the category "Therapist Incompetent". Results of the survey showed that the categories most frequently present in the accounts were 'Therapist Threatened', followed by 'Therapist Puzzled' and 'Therapist Incompetent'. The categories least

frequently represented were (in descending order) 'Therapist Stuck', 'Therapist's Personal Issues' and 'Moral/Ethical Dilemma'.

While this study succeeded in confirming that the difficulty taxonomy is capable of classifying the content of accounts drawn from a range of therapists reliably, it failed to yield solid results in other areas. Notably, there were no substantial associations between difficulty categories and therapist or situational variables. The only material and stable association between a difficulty category and a patient variable consisted of a correlation of r = .40 between the presence of the category 'Therapist Threatened' in an account and the description 'borderline' being applied by the therapist to the patient with whom the difficulty occurred. The experience of particular difficulties was also not predictive of therapists' specific efforts to deal with them, as there were no substantial correlations between difficulty categories and coping strategies (which had also been rated from the accounts on the basis of a taxonomy developed in parallel to the difficulty categories (Davis et al., 1989)). Finally, intercorrelations between difficulty categories - calculated separately for each of the three situations supplied by respondents in order to avoid dependencies between the datapoints - were generally either low (less than .3) or did not replicate across situations. Consequently, principal component analyses did not yield interpretable factor structures.

At this point the results seemed to indicate that therapist difficulties, as construed by the taxonomy, are discrete phenomena which can be clearly delineated and reliably identified, but which are unrelated to the features of the setting in which they occur, largely independent of the characteristics of the patients with whom they are experienced, unconnected to the attributes of the therapists reporting them, not meaningfully associated with each other, and not predictive of the particular ways in which therapists try to cope with them. As a basis for further investigation this

situation appeared less than promising. There were, however, two exceptions to the overall picture. The first stemmed from the study establishing the preliminary taxonomy categories, whereas the second emerged from a re-analysis of the survey data prompted by the results of a subsequent investigation discussed in more detail under 2.2.1.2.2 below. Each will be considered in turn.

One of the questions investigated in the preliminary study reported by Davis et al. (1987) concerned the representativeness of a particular combination of difficulties for a particular therapist. If therapists can indeed be differentiated according to the types of difficulties they experience, then one would expect 'profiles' of taxonomy categories for difficulty accounts to vary more between therapists than within therapists. As six investigators had each contributed five accounts of difficulties for the study, and as the taxonomy comprised nine categories at the time (with 'Painful Reality' and 'Therapist's Ethical Dilemma' still combined in a single category), a 6 x 9 (therapist x category) ANOVA of mean presence-absence ratings (averaged over seven raters) with five situations per cell was carried out in order to examine this question. The results showed a significant ($p < .05$) interaction between categories and therapists, demonstrating that category profiles did actually vary more between than within therapists, thus providing some evidence for the presence of idiosyncratic difficulty patterns among different therapists. Indications for the validity of such patterns could be found by matching some of the category profiles to the perceptions which members of the collaborative research group had developed about each other's personal characteristics. One researcher's propensity to take on difficult and vulnerable clients for 'heroic' therapies was matched by an elevation on scores for the category 'Therapist Damaging', another's pervasive concern over the maintenance of firm therapeutic boundaries was reflected in raised 'Therapist Threatened' scores, a

third collaborator's contemporaneous domestic crisis cast light on comparatively high scores for the category 'Therapist's Personal Issues'. At the time, the group did not focus on one distinguishing feature among these three examples which is of particular concern to the current study, i.e. the relatively enduring nature of the first two instances, compared with the more temporary quality of the third. However, the identification of therapists' idiosyncratic difficulty profiles - which might also provide an operationalisation of the psychodynamic concept of countertransference - was recognised as having "significant diagnostic value for [therapists'] training and supervision" and was identified as an important area for further investigation (Davis et al., 1987, p. 117).

The second exception to the general dearth of meaningful associations with and between difficulty categories was discovered (and eventually reported (Schröder, 1997)) by the present author in a re-analysis of the survey data in the light of information obtained from the 'International Study of the Professional Development of Psychotherapists' (ISPDP) (see 2.2.1.2.2). Analysis of data from that study had suggested that a stable and interpretable factor structure could be derived from responses to a 20-item questionnaire, constructed on the basis of the taxonomy categories, enquiring about the frequencies with which therapists experience each of the difficulties. The results, based on a data set of 2376 respondents, suggested a three factor solution in which experiences of difficulties tended to co-occur according to where therapists located their source: either in themselves (which earned the factor the shorthand-title 'Bad Therapist '), or in the client ('Bad Patient'), or in external circumstances ('Bad World'). The first of these factors showed associations with indices of therapist development, in that difficulties of the 'Bad Therapist' variety were attenuated by increasing practical and professional experience (Davis, Davis,

Schröder, Orlinsky, & Howard, 1992; Davis, Schröder, Davis, & Orlinsky, 1996). Responses to the Difficulties Scale in the ISPDP are elicited under the heading "About Your General Experience Of Difficulties In Therapeutic Practice" by the prompt "Currently, how often do you feel...?" which is designed to capture a global impression of therapists' experiences of difficulties in their entire practice. The equivalent strategy for the survey data would be to aggregate therapists' responses from all three accounts, thereby creating the kind of category profiles described above. Adopting such a strategy, the results of Principal Components Analyses with subsequent Varimax rotations derived from the correlation matrix of mean presence-absence ratings averaged over three situations for all ten categories did indeed yield interpretable factor structures. Using the criterion of retaining factors with eigenvalues greater than one would have suggested a four factor solution. However, the scree plot indicated a three factor solution which also comes closest to the ISPDP results and which is therefore reported here.

The three factors, which account together for 52% of the total variance, are made up as follows *(categories were assigned to a factor if they loaded above .5 on it and if the loading was at least .15 higher than that on any other factor; figures in square brackets denote internal consistencies of the resulting factor scales expressed as Cronbach's alphas calculated on the basis of N = 96 respondents):*

Factor 1 [α = .62] ("Therapist Bungling") comprising the categories of "Therapist Incompetent", "Therapist Damaging" and "Therapist's Personal Issues".

Factor 2 [α = .56] ("Difficult Patient") comprising the categories of "Therapist Stuck", "Therapist Out Of Rapport" and "Therapist Thwarted".

Factor 3 [α = .41] ("Therapist Pained but not Threatened") comprising

"Painful Reality" and - loading negatively - "Therapist Threatened".

These factors appear to concur roughly with those derived from the Difficulty Scale in the ISPDP questionnaire, the conceptually weakest correspondence being that between Factor 3 and the "Bad World" factor.

Taking both of the above findings together, they seem to form a basis for concluding that further exploration of therapists' idiosyncratic patterns of difficulties might prove fruitful for a better understanding of both therapy process and therapist development.

1.3 Development of Research Questions

Apart from the findings reported above - which implied the heuristic value of the concept of distinctive difficulty patterns for further explorations - there were two other, less formally derived, apperceptions which contributed to the formulation of the research questions for this thesis.

The first originated from an attempt to chronicle session-by-session difficulty experiences in the course of a few therapies which I had been conducting myself. It soon became apparent that some difficulties were impermanent, occurring only in one or two sessions, while others featured repeatedly in the notes I was making over the course of therapy. An example of the former would be one client's insistence in a first session that her problems were entirely due to somatic causes, which the referring doctor was said to have missed despite repeated investigations, while an example of the latter would be the observation that a particular patient often seemed to be disengaged and hard to reach until the last ten minutes of his weekly sessions.

The second flowed from the experience of supervising psychodynamic psychotherapists in training over long periods of time (three to five years), while periodically reflecting and reporting on their struggles. Here it became clearer that

some difficulties I was hearing about were enduring over the timespan of training - for instance a limited capacity to relate empathically to grandiose ideas of entitlement in clients with a very fragile sense of self. Other difficulties seemed to be of a transient nature, disappearing in the passage of a few weeks or even within a single session - for example a feeling of irritation with a patient's intrusive personal questions. In between these two extremes, there seemed to be some difficulties which would persist for a substantial interval and then diminish in importance - for instance a reluctance to address feelings related to breaks in therapy, for fear of overvaluing one's importance in a client's life.

The initial working titles chosen for these kinds of difficulty experiences were 'paradigmatic' for the enduring difficulties (expressing the conjecture that these were in some sense typical for the therapist experiencing them or the therapy in which they occurred), 'transient' for the short-lived problems and 'episodic' for the intermediate variety. The attempt, in a further step, to link this proposed spectrum to developmental concepts produced the idea, first proposed in a conference paper (Davis, Davis, Davidson, & Schröder, 1990), that it would also map on to three different models of development:

a) continuous, incremental change (as for instance in the acquisition of taught skills) for TRANSIENT difficulties;

b) discontinuous, phased change (as for instance in epigenetic models such as Erikson's stages of psychosocial development (Erikson, 1980)) for EPISODIC difficulties; and

c) progressive adaptation to largely invariant structures (as for instance in maturational models such as Vaillant's (1976) hierarchy of defence mechanisms) for PARADIGMATIC difficulties.

Establishing that such distinctions can reliably be made would have obvious consequences for tailoring supervisory responses according to the domains to which difficulties belong. The first research question was therefore formulated as

Question 1: Is it possible to distinguish among TRANSIENT, EPISODIC and PARADIGMATIC difficulties such that these can be construed as lying on a continuum of 'pervasiveness'?

The logical next step in the enquiry was the attempt to find defining criteria for the continuum postulated above. Drawing mainly on my personal experience both as a therapist (with a retrospective understanding of my own training and development) and as a clinical supervisor (observing therapists in training in their struggles and charting their evolving professional competence) the following dimensions crystallised:

1) 'Familiarity': It seemed likely that paradigmatic difficulties, having persisted for longer periods, would be well- known to therapists, while transient difficulties would have been in evidence for only a short time.

2) 'Depth of impact': It seemed likely that paradigmatic difficulties, because of their origins in therapists' personal attributes and because of their enduring nature, would affect therapists more deeply than transient difficulties.

3) 'Duration': By definition, transient difficulties are short-lived, while paradigmatic difficulties are not readily resolved and therefore likely to continue for longer.

The second research question was therefore formulated as follows:

Question 2: Are the dimensions of 'familiarity', 'depth of impact' and 'duration' associated with the spectrum of difficulties in such a way that high levels of these qualities denote the 'paradigmatic' pole of the continuum, while low levels denote the 'transient' pole?

It was further postulated on the basis of the developmental concepts described above that transient difficulties would be related to therapists' skills deficits while paradigmatic difficulties would be associated with therapists' enduring characteristics[5]. Episodic difficulties might then correspond to specific phases in therapists' professional development (Schröder & Davis, 1993). This line of argument led to the formulation of the third research question:

Question 3: Can it be demonstrated that paradigmatic difficulties, as defined by the parameters outlined in Question 2, are associated with therapists' enduring personal characteristics, whereas transient difficulties are associated with gaps in the skills, knowledge or experience of the therapist reporting them?

Before going on to describe how these questions were operationalised and investigated in three separate studies, based on two consecutive surveys, an attempt will be made to locate the field of this enquiry in the relevant literature.

[5] The idea was not of an isomorphic mapping of particular personality 'traits' on to particular difficulties, but rather of the interaction between therapists' individual characteristics and particular features of the situations or the patients encountered creating idiosyncratic difficulties.

2. Chapter Two: Location in the Literature

This heading has been chosen to indicate that there are a number of areas in the literature on psychotherapy research which have connections to the investigations in this thesis without providing a theoretical basis for them. No attempt will be made to review all of these areas in detail; however, reference will be made to those studies which have a bearing on the present topic. Where specific measures used in the second Therapist Difficulty Questionnaire (TDQ2) (see 5.2.1.2 below) are derived from a particular theoretical framework, this will be reviewed in general terms here, while the measures themselves will be reviewed in Chapter Five.

2.1 Other Direct Investigations of Therapist Difficulties

Apart from the work of the Warwick Group, there are three previous studies which can be seen to have addressed therapist difficulties directly and which will be reviewed in turn. While none of them makes a direct contribution to the understanding sought in this investigation, at least two of them yield some interesting points of comparison and stimuli for further thought.

2.1.1 Critical Incidents

One previous investigation (Plutchik, Conte, & Karasu, 1994), which seeks to study events creating difficulties for psychotherapists, takes its point of departure from critical-incident technique (Flanagan, 1954). The authors, who define critical incidents as "behavioral events that have a special relation to some outcome" (p. 75), assembled a list of 52 "unusual or infrequent events that create difficult problems for the therapist" (p. 77) by pooling experiences from a group of seven experienced clinicians. The items on the list consist of behavioural descriptions such *as "patient shows up very late"* or *"patient is silent for three minutes or more"*, although some

incidents are not as tightly operationalised, for instance *"patient insults therapist"* or *"patient rejects therapist"*. Fifty items describe patient behaviours, the remaining two focus on the actions of "a member of the patient's family".

The complete list was presented in survey format to a group of 21 experienced psychiatrists (mean length of post-M.D. experience 12 years, range 2 - 25 years) and to a group of 10 inexperienced psychiatrists (third-year residents) who rated items for 'importance' - defined as "potential for affecting the future course of therapy" - and 'frequency' - how often an incident had occurred in the practice of the rater. The three most important items identified by mean ratings of the 'experienced' group were *"patient threatens therapist physically"*, *"patient threatens suicide"* and (perhaps unexpectedly) *"patient has seen a second therapist and reveals it"*. The most frequently experienced incident in the same group was *"patient reports a major trauma in his (her) life (e.g. death of a spouse or parent)"*, while the least frequently experienced was *"patient threatens therapist physically"*.

A different group of 6 clinicians was involved in a qualitative study attempting to categorise incidents by content area. Four distinct categories were labelled respectively as (a) "threat of harm" (3 items); (b) "criticism of the therapist or of the therapy" (10 items); (c) "occurrence of a major life event" (5 items) and (d) "attempt by the patient to seek friendship or achieve seduction" (12 items). Reliabilities for assigning items to categories ranged from quite modest (Kappa = .33) for category (d) to good (Kappa = .72) for category (a). The remaining 22 items were assigned to a "Miscellaneous" category for which reliability was slight (Kappa = .21). Finally, a comparison of the ratings of the 'experienced' and the 'inexperienced' groups showed that the latter judged most incidents as less important but more frequent.

There are some methodological and conceptual issues which seriously limit the value of the findings of the above study. Most notably, the emphasis on descriptions of patient behaviour, rather than on the meanings these behaviours have for the therapist, gives rise to problems. On the one hand, incidents can be overly specific; for instance *"patient tries to kiss therapist"* begs the question why other items such as 'patient tries to hug therapist' were not included. On the other hand, more generalised descriptions, such as *"patient challenges therapist's abilities"* are open to different interpretations as to their effect on the therapist. The authors acknowledge as much when they report that the item *"Patient says 'I hate you' to therapist"* was taken by one half of judges to imply criticism of the therapist, whereas the other half understood it to be an attempt to seek friendship. Even accepting the behavioural focus of the assembled incidents, it seems unnecessarily constraining to restrict difficulties to those originated by the patient (or a member of the patient's family) without allowing for therapist behaviours which might give rise to difficult incidents.

The qualitative study not only suffers from low reliabilities in the assignment to those categories containing more than three items, but also has a large heterogeneous 'all purpose' classification which breaches the principle of 'internal cohesiveness' held to be important for qualitative studies using category assignments (Fonagy & Moran, 1993).

Despite these shortcomings there are some features which overlap with the concerns of the current inquiry. Notably, the 'frequency' category is similar to the 'familiarity' concept, presumed to differentiate transient from paradigmatic difficulties. As mean frequencies for both the 'frequency' and the 'importance' categories had been provided in the published paper, it was possible to carry out a further data analysis, investigating the relationship between the concepts. The results show that overall

there is a moderate but highly significant association between the two concepts (r = -.45; $p < .001$), which is most pronounced for the "threat of harm" category (a) (r = -.86; n.s.), and the 'boundary infringement' category (d) (r = -.84; $p < .001$), somewhat attenuated for the "life event" category (c) (r = -.55; n.s.), and inconspicuous for the "criticism" category (b) (r = -.31; n.s.). This suggests that 'familiar' events, which should be located towards the 'paradigmatic' end of the spectrum proposed above, are judged least likely to have a significant effect on the future development of the therapy. This is most pronounced when they are of a nature suggesting high emotional impact on the therapist, which is again indicative of paradigmatic difficulties. These results suggest that for familiar events, impact on therapy (for better or worse) and impact on the therapist are inversely related. One might say, that faced with a difficulty of the paradigmatic type, therapists have little conviction that the course of therapy might alter in its wake - they are affected but do not anticipate change. This would be consistent with a view that these familiar, high-impact events are mainly related to therapists' internal worlds.

2.1.2 Difficult Moments...

Another series of studies, reported under the heading of 'Difficult moments in difficult sessions with difficult patients' (Freedman, 1992a), sets out to explore the structure of a particular process sequence which is held to be central to the therapeutic action of psychoanalysis. The postulated steps in this sequence are

1) 'the patient's transference' (within the framework adopted in this thesis this would be the antecedent of a therapist difficulty);
2) 'the evocation of the analyst's countertransference' (the difficulty experience itself);

3) 'the use or reworking of the countertransference prior to the analyst's interpretation' (a form of coping with the difficulty which in the terms of Davis et al. (1989) would come under the heading of 'Therapist turns to Self'); and finally

4) 'the reorganisation of the patient's psychic life' (the consequence for the patient of the therapist's successful coping).

The second step of this sequence is closest to the subject matter of this thesis. Rather than relying on therapists' self-report, Freedman, Lavender, & Berger (1992) focus on kinesic enactments, that is particular bodily movements which are held to correspond closely to emotional states in the therapist. One of these, termed 'shielding', which involves the bringing together of both hands in front of the body accompanied by slight finger movements, serves as a marker of experiences of difficulties. For process studies which utilise direct observation of therapist behaviours, this marker provides a potential form of external validation of the difficulty concept if used in conjunction with retrospective verbal report.

An example of such a retrospective account of a therapist's experience during a difficult moment, reported by Freedman, Lavender, & Hoffennberg (1992), allows for comparisons with the kind of description typically related by respondents to the survey of British therapists outlined in 1.2 above. Although some of the content of that narrative can be classified in terms of the difficulty taxonomy described above (with the categories of T-out-of-Rapport and T-Thwarted most prominently in evidence), there are also clear differences. Apart from being a much more fine-grained account of a whole sequence of experiences, there is also a stronger focus on bodily sensations in Freedman's example, in line with the preoccupations of the study in which it occurs. Most importantly, the account is based on general recall of in-

session experiences rather than on the cued recall of specific difficulties which elicited the descriptions in the studies constituting this thesis. However, the finding that observable body language and subsequent verbal report of a difficulty experience can be linked together, lends weight to the notion that therapist difficulties play an important role in psychotherapy process.

The feature of Freedman's study which has been most stimulating for the current inquiry is that of differentiating 'difficult' from 'not so difficult' patients. Originally based on a subjective judgement by therapists and their supervisors, selecting one representative of each category from a caseload of in-patients in a therapeutic community (Freedman, 1992b), the distinction was subsequently validated by differential scores on a composite measure of 'symbolized transference' (Berzofsky, Freedman, Essig, & Margolis, 1992). The ramifications of this for the design of TDQ2 are discussed below in 5.2.1.1.1

2.1.3 Conscious Countertransference

Torres (1983), in an unpublished PhD dissertation, employs an approach which has a number of similarities with that taken by the Warwick Group. Referring to, but departing from, previous empirical investigations of the psychoanalytic concept of countertransference, which traditionally have understood this to be an unconscious occurrence which can only be tapped indirectly by studying its behavioural concomitants, she sets out to focus in her work on the conscious, experiential rather than the behavioural aspects of this phenomenon. Not unlike the approach eventually adopted in this thesis, Torres combines a phenomenological investigation, using data derived from structured interviews, with an empirical investigation, using data obtained with a pre-existing measure (the Impact Message Inventory (IMI) (Kiesler et

al., 1976)). Her study is, however, restricted to one particular kind of therapist - doctoral psychology students in their second year of clinical experience, and to one particular kind of therapist experience - their reactions to patients' expressions of anger. There are, however, three interesting features in her work which have some connection to the current enquiry:

a) The attempt to capture cues which could alert therapists to the imminence of conscious countertransference reactions;

b) the conjunction of conscious countertransference reactions in therapy with some measure of reactions in private relationships; and

c) the association between therapists' reactions and their ultimate choice of intervention.

Of Torres' results, the most pertinent one is that therapists with the least patient hours tended to view their patients as significantly more aggressive than therapists with the highest patient hours. This could suggest that perceived attacks from patients should be over-represented in transient difficulties (which are predicated on lack of experience). Given that even the most experienced therapists in her study were all but novices, generalisations to a more broadly sampled group of therapists are, however, at best tenuous.

2.1.4 Conclusions for this Enquiry

Having reviewed three previous studies which sought to investigate an aspect of therapist difficulties directly, the question arises as to what can be learned from them for the present enquiry. Some points which stimulated thinking, such as the distinction between 'difficult' and 'not so difficult' patients, have already been mentioned. Some potential links have also been pointed out. But the most important

information obtained is negative: There is an absence of studies in an area which is of such obvious practical consequence to clinicians. Those few studies which have been undertaken are either highly circumscribed or incidental. There is no attempt so far to formulate a theoretical base for understanding the described phenomena[6] and no attempt at a more broadly-based data collection. Investigating therapists' experiences of difficulties in their practice means moving into largely uncharted territory and provides a challenge not only to give therapists a voice in relating their own perceptions, but also to begin to construct a frame of reference which might help to make sense of those observations.

[6] Freedman's (1992a) theorising was about the mechanism of what, in a different frame, one might call 'projective identification' rather than about therapist difficulties.

2.2 Location in the Literature - Areas Pertinent to this Enquiry

2.2.1 Psychotherapist Research

Despite some more recent attention, for instance Strupp (1995), there is little research to date on the psychotherapist other than as an independent variable influencing outcome. Apart from some general descriptive studies, the specific area of the professional development of psychotherapists has been subject to investigation, as described below.

2.2.1.1 General Studies

Norcross & Wogan (1983), in one of the earliest studies focusing on the psychotherapist in his/her own right, reported results from a survey among North American therapists, describing their demographic characteristics, clientele and theoretical persuasions. Influences of the experience of conducting psychotherapy on the therapist have been studied by Farber (1983a) and, with particular reference to the role of personal therapy, by Norcross & Prochaska (1986) and more recently by Pope & Tabachnik (1994) and Botermans, Orlinsky, & Wiseman (1998). Stress arising from client contact has been a prominent interest, (Deutsch, 1984; Rodolfa, Kraft, & Reilly, 1988), together with coping strategies preventing burnout (Medeiros & Prochaska, 1988).

2.2.1.2 Therapist Development

There are a number of theoretical models of therapist and counsellor development described in the literature, for instance by Hogan (1964), often in relation to supervision (e.g. Loganbill, Hardy, & Delworth, 1982), and generally following a

stage format (Stoltenberg & Delworth, 1987). Two salient empirical studies are described below, one mainly qualitative, the other quantitative.

2.2.1.2.1 The Minnesota Study

Skovholt & Rønnestad (1992, 1995) interviewed 100 counselling and psychotherapy professionals practising in Minnesota, representing a wide span of experience ranging from novices (first year of graduate school) to senior practitioners (40 years of post-qualification experience). After devising stage descriptions, they re-interviewed 60 informants to check the authenticity of their portrayals and proceeded to abstract themes and construct a theoretical model. A central feature of the model is the process of 'continuous professional reflection', enabling individuals to assimilate or accommodate to negative experiences and to stay on a 'developmental track'. This is contrasted with a process of selection or distortion leading to premature closure or pseudo-development on a 'stagnation track'. Although not commenting on difficulties in practice directly, the model has clear implications for the contribution that their successful handling can make to therapist development.

2.2.1.2.2 The ISPDP

The International Study of the Professional Development of Psychotherapists (ISPDP) is a multi-centre collaborative investigation of aspects of therapeutic practice and development. Its main instrument, the Common Core Questionnaire (CCQ), has so far been translated into 17 languages and (as of June 1998) more than 4000 therapists have completed it. Although a longitudinal questionnaire has been developed and is in use in some countries, the majority of developmental information is derived from cross-sectional studies. However, respondents are also asked to evaluate their own development to date retrospectively.

Therapist difficulties and coping strategies are assessed by two scales inquiring about the frequency with which they occur in a respondent's current practice. Rank-orderings of difficulty items and the factor structure of the scale are both remarkably stable across national subsamples[7] (Schröder, 1997). As outlined in 1.2 above, difficulties located within the therapist are attenuated by professional development. In this thesis, such difficulties are further sub-divided into transient and paradigmatic types, with only the former thought to be influenced by improvements in therapists' competence levels .

2.2.2 Psychotherapy Research - Therapist Variables

There are numerous studies investigating characteristics of psychotherapists, usually as independent variables related to therapy outcome. Beutler, Machado, & Neufeldt (1994), in their review of the extensive literature, seek to organise the field along the dimensions of 'objective' vs. 'subjective' characteristics and 'cross-situational' vs. 'therapy-specific' states. 'Therapist distress' is listed amongst the subjective cross-situational states as a topic which has received some attention; and is closest to the concerns of this thesis. However, therapist distress is usually not understood in terms of specific situations but rather in terms of general severity of symptoms or global disturbance, either arising from external factors, such as stressful life-events (Greenspan & Kulish, 1985), or intrinsic factors, such as work-stress (Farber, 1983a) or burnout (Beck, 1988 as cited in Beutler et al., 1994). The notion that the stimulation of therapists' internal conflicts may affect in-session behaviours has received scant attention (e.g. Cutler, 1958) and overlaps with the literature on countertransference.

---

[7] With the exception of the South Korean sample.

2.2.3 Psychotherapy Research - Process Studies

Psychotherapy process research has been extensively studied, however, literature on the area of direct relevance to this thesis - therapists' experience of process - is sparse. Apart from an early study by Snyder (1961), the most salient work is that by Orlinsky & Howard (1975, 1977). Their work with the Therapy Session Report (TSR) led to the identification of two factors, 'patient distress'(PD) and 'therapist effectiveness'(TE), allowing for a typology of sessions based on high and low levels of each - 'smooth sailing' (high TE, low PD), 'heavy going' (high TE, high PD), 'coasting' (low TE, low PD), and 'foundering' (low TE, high PD). Low therapist effectiveness certainly constitute a component of felt difficulties, especially those of the transient type. Paradigmatic difficulties, however, seem to call for a (situation-specific) therapist distress category not covered by the questions in the TSR.

2.2.4 The 'Difficult Patient'

Contact with troublesome clients as a source of stress (Farber & Heifetz, 1981) has already been commented on above, and the notion that therapists can identify difficult patients among their caseloads was confirmed by Freedman (1992a) (see 2.1.2 above). Farber (1983b) identified two dimensions causing patients to be perceived as difficult by their therapists - psychopathology (especially threat of suicide) and resistances to the treatment process. Robbins, Beck, Mueller, & Mizener (1988) arrived at a more differentiated picture, identifying five factors, and concluding that patients denying the competence and authority of clinicians are most likely to be experienced as difficult. Fremont & Anderson (1988), using therapist annoyance as a marker, found that imposing and attacking clients were most likely to elicit negative reactions from their therapists. Difficult patients are not confined to psychotherapeutic or psychiatric settings. A recent study by Mathers, Jones, &

Hannay (1995) investigated 'heartsink patients' in general medical practice and found the frequency of their occurrence to be linked to skills, attitudes, and general stress level of their doctors. One of the studies in this thesis (see 5. below) also yields findings on the factors leading therapists to perceive their patients as difficult.

2.2.5 The Interpersonal Perspective

This particular theoretical tradition is included in this review for two reasons:

a) It provides one theoretical framework to account for the formulation of research question 3 (see 1.3 above) and for the eventual definition of paradigmatic difficulties (see 4.2.1.2 below), giving rise to the notion that pervasive problematic interactional patterns - which are characteristic for particular therapists - are activated during the experience of paradigmatic difficulties;

b) one of its theoretical constructs, the interpersonal circumplex, has been the subject of extensive empirical investigation and forms the basis for two of the measures used in TDQ2 (see 5.2.1.1.5.2 and 5.3.4.4 below).

In clinical practice, there is considerable convergence between interpersonal approaches and the British Object Relations tradition and both theoretical frameworks have influenced my own thinking and practice.

2.2.5.1 Historical Development

The roots of contemporary interpersonal theory can be traced to Sullivan (1953) who saw human behaviour as dependent on its interpersonal context, and personality as the internalisation of interactional patterns. Subsequently, writers such as Carson (1969) and Kiesler (1973) developed the idea that psychopathology can be understood in terms of inflexible maladaptive behavioural patterns which 'pull for' (evoke) complementary responses from others, leading to predictable but self-defeating

interactional cycles. Leary (1957) introduced the Interpersonal Circle (IPC), organised around the two orthogonal axes of affiliation and dominance, based on the idea that in human interaction two major issues are constantly negotiated - how friendly the participants will act towards each other and who is in charge of the interaction. Circumplex models (Gurtman, 1997) of interactional behaviour have since been a major focus of interpersonal theory and research.

2.2.5.2 The SASB Model

'Structural Analysis of Social Behavior' (SASB) is a development of the IPC introduced by Benjamin (1974) taking up Schaefer's (1965) version of the circumplex for parenting behaviours and complementing it with circumplex maps for childlike behaviours and impact on self-concept (Benjamin, 1996b), to create "three interpersonal surfaces versus one" (Benjamin, 1994b, p.278). This model allows for the integration of the traditional arrangement of the IPC with Schaefer's views, who used 'grant autonomy' instead of 'submit' in order to define the opposite pole of 'dominate' on the vertical axis. The three 'surfaces' (circumplexes) are labelled surface 1, 'transitive' (parent-like transactions, active, focus on other); surface 2, 'intransitive' (child-like transactions, reactive, focus on self); and surface 3, 'introject' (internalised transactions, intrapsychic, inward focus).

Originally, SASB was developed as a coding system for interpersonal behaviour and its initial applications in the field of psychotherapy research were in the area of family therapy (Benjamin, 1979; Benjamin, Foster, Giat-Roberto, & Estroff, 1986). Subsequently, SASB codings were also used as a process measure in individual psychotherapy (Benjamin, 1982; Henry, Schacht, & Strupp, 1986). Codings involve three decisions per behavioural unit:

a) The focus of the transaction from the point of view of the subject (referent), leading to a decision about the appropriate surface;

b) the degree of affiliation represented by the transaction, rated on a 19-point scale ranging from -9 (extremely hostile) to +9 (extremely friendly); and

c) the degree of autonomy represented by the transaction, also rated on a 19-point scale, ranging from -9 to +9 (meanings of these extremes vary according to surface, labels for the positive poles are 'endorse freedom'[8], 'freely come and go', 'happy-go-lucky' for surfaces 1,2, and 3 respectively, while the corresponding negative poles are labelled 'manage, control', 'yield, submit, give in', and 'control, manage self' (Hartley, 1991).

Locations of coding units on the appropriate surface are therefore determined by polar co-ordinates derived from the affiliation and autonomy ratings. The full SASB model consequently has 36 points (eight per quadrant plus the four extreme poles) per surface. There are two lower levels of complexity: The simplest is the quadrant model which collapses ratings into composite quadrant scores, labelled (from the top in counter-clockwise order for surfaces 1/2/3) 'invoke hostile autonomy'/'take hostile autonomy'/'reject self', 'hostile power'/'hostile comply'/'oppress self', 'friendly influence'/'friendly accept'/'manage, cultivate self', and 'encourage friendly autonomy'/ 'enjoy friendly autonomy'/'accept, enjoy self' (Benjamin, 1984). Between these two renderings of the model lies the octant version which collapses the points of the full version into eight 'clusters'. It is thus most similar to other circumplex models which are often also subdivided into eight segments (e.g. Birtchnell 1994; Kiesler, Schmidt, & Wagner 1997). As this version also forms the basis for the SASB questionnaire used in TDQ2, cluster labels are mentioned in detail

---

[8] This pole, later also called 'emancipate' (Benjamin, 1996a) corresponds to Schaefer's (1965) 'grant autonomy'.

below (see 5.3.5.1). SASB questionnaires (called INTREX, for reasons not explained in the literature) have been developed as a self-report complement to the coding system (Benjamin, 1984) and represent either the full version ('Long Form questionnaire') or the cluster version ('Short Form questionnaire').

SASB operationalises a number of interpersonal principles, notably interpersonal similarity (identical topological positions adopted by two referents on matching surfaces), interpersonal complementarity (identical topological positions adopted by two referents on two different surfaces), and antithesis (diametrically opposite topological positions adopted by two referents on two different surfaces).

SASB has given rise to a wide variety of applications, ranging from a detailed study of parallel process in supervision (Alpher, 1991) to investigations of the characteristics of groups as diverse as care-givers (France & Alpher, 1995) and antisocial alcoholics (Ichiyama, Zucker, Fitzgerald, & Bingham, 1996). It has been proposed as a theoretical underpinning to DSM III (Benjamin, 1987) and DSM IV (Benjamin, 1996a) categories; has been suggested as a possible theoretical component of parallel distributed processing models of cognitive schemata (Benjamin & Friedrich, 1991); and is explicitly understood by Benjamin (1986, 1993, 1994a) as capable of operationalising concepts from Object Relations Theory.

2.2.6 Transference and Countertransference

The concept of transference is included in this section because it has close links to a particular way of viewing the expression of internal conflicts through interaction (the 'core conflictual relationship theme' (CCRT)) which formed the basis for one of the measures in the third study in this thesis (see 5.2.1.1.7.5 below). Countertransference,

as mentioned in the introduction above, has a connection to the notion of paradigmatic difficulties.

2.2.6.1 Transference

2.2.6.1.1 Theoretical Concept

Transference, the occurrence of thoughts, feelings and perceptions belonging to past relationships in the context of the current therapeutic relationship, is one of the distinguishing characteristics of psychodynamic therapies. Originally thought of as an impediment to treatment (Rycroft, 1972), its occurrence and understanding was subsequently seen as a central agent of therapeutic change (Sandler, Dare, & Holder, 1973). Although criticised from a person-centred perspective (Shlien, 1987), it has become widely accepted among therapists from different theoretical persuasions (e.g. Basch, 1980). Although most readily visible in the consulting room, it can be thought of as a general explanatory principle for interpersonal perceptual distortions.

2.2.6.1.2 The CCRT Method

Originally developed as a general measure of relationship patterns (Luborsky & Crits-Cristoph, 1989) in conjunction with the helping alliance scale (Luborsky, 1976), the CCRT method was later employed to examine a number of Freud's observations about the transference (Luborsky et al., 1985; Luborsky, Crits-Christoph, & Mellon, 1986). Narratives of an individual's interactions with others are frequently told in psychotherapy. Such 'relationship episodes' are organised by the CCRT method into three steps - the (often unconscious) wishes, needs, or intentions directed by an individual towards important others; the responses from these others as perceived by the individual (R-O); and, finally, the responses of the self (R-S), experienced by the individual as a reaction to the responses from others. On the basis of these three steps,

clinicians (or raters in a research context) infer interactional sequences which, if found repeatedly, are held to relate to core conflicts. Their stability suggests that they are in part based on perceptual distortions and hence to transference. Reliability and validity of CCRT judgements have been established in a series of studies (Luborsky, Crits-Christoph, Mintz, & Auerbach, 1988).

In reviewing possible future directions for the empirical investigation of *countertransference* (see next section), Crits-Christoph, Barber, Baranackie, & Cooper (1993) single out the possible match between the therapist's and the patient's CCRT as a particularly promising area of study. A similar idea has been used in the third study of this thesis (see 5.2.1.1.7.5 below).

2.2.6.2 Countertransference

2.2.6.2.1 Theoretical Concept

There is a wide divergence of views on the scope of this concept (Sandler et al., 1973), ranging from a narrow definition - countertransference as the therapist's reaction to the patient's transference - to an encompassing view - countertransference as the sum total of the therapist's emotional reactions during treatment. Again, originally viewed as an impediment to treatment, countertransference has now gained a central place in psychodynamic theory and therapeutic practice, and has produced a prolific clinical literature. Some of this, for instance investigations of therapeutic impasse (Elkind, 1992), or vicarious traumatisation (Pearlman & Saakvitne, 1995), has a direct bearing on the topic of this thesis.

2.2.6.2.2 Empirical Studies

Singer & Luborsky (1977) in their comparative review of clinical and empirical literature note the sparsity of empirical studies. This situation has not altered greatly

in the intervening 20 years. Empirical studies, whether early (Fiedler, 1951; Cutler, 1958; Bandura, Lipsher, & Miller, 1960), or more recent (Singer, Sincoff, & Kolligian, 1989), typically focus on therapists' perceptual distortions or interruptions of their therapeutic performance, rather than on their experience. While the notion of similarities between therapists and patients giving rise to conflicts has been fruitful for the current study, no other features of direct relevance emerged from these studies.

3. Chapter Three: Survey 1 - The First Study

This chapter describes the first attempt at operationalising and validating the proposed transience-paradigmatic difficulties spectrum through data collected in a survey of British psychotherapists with the first "Therapists' Experiences of Difficulties in Practice" questionnaire (TDQ1).

3.1 Rationale

Before being able to study their correlates and their relationships with each other, one needs to be able to identify transient and paradigmatic difficulty types reliably. Criteria for classifying difficulties had to be found which were independent of constructs such as 'skills deficits' or 'enduring personal characteristics' to avoid confounding later investigations. Although 'familiarity', 'impact' and 'duration' were at this point only hypothetical characteristics of the transient-paradigmatic spectrum, one could nevertheless assume that a substantial number of difficulties would be associated with the experience of high levels of all three attributes, allowing them to be interpreted as 'paradigmatic, whereas others would be associated with low levels, allowing them to be classified as 'transient'. If such constellations should indeed occur, difficulty types and experiential attributes would in effect validate each other.

The initial hypothesis could therefore formulated as follows:

- High levels of 'familiarity', 'depth of emotional impact', and 'duration' measured in relation to self-reported difficulties will occur simultaneously, as will low levels.

3.2 Method

As Everitt (1996) observes, "survey methods are based on the simple discovery 'that asking questions is a remarkably efficient way to obtain information from and about people'" (p. 8). Asking therapists about their difficulties and associated experiences in a survey format seemed the most suitable way of eliciting the information needed to investigate the above hypothesis. Interviews were not considered at this stage, partly because of the time needed to conduct (and transcribe) a sufficient number, but also because the anonymity of a postal survey was thought to enhance honest reporting of difficulties and therefore the richness of data collected. The instrument for the survey was a questionnaire (TDQ1), designed to elicit self-reports of difficulties, associated experiences, and a number of variables relating to respondents and the patients with whom difficulties were experienced.

3.2.1 Instrumentation

3.2.1.1 New Measures

As a first step in the investigation, it was necessary to obtain reliable measures of the three postulated concepts (Familiarity, Depth of Impact and Duration) which were thought to define the Transient - Paradigmatic spectrum of therapist's experience of difficulties. In the absence of suitable pre-existing instruments, three new measures - two scales and one index - were developed. Their incorporation in the survey allowed for the empirical investigation of their psychometric properties and their associations with each other.

3.2.1.1.1 Scale Development

Measures for 'Familiarity' and 'Impact' were constructed as scales assessing the accuracy of each concept in describing the experience of a difficulty. These scales were composed by assembling items designed to capture different facets of each concept rather than to define various levels of a single 'prototypical' item semantically. The same strategy was adopted for the development of both scales. It consisted of

a) expounding the concept by defining it and distinguishing it from related ideas;

b) expanding the concept by differentiating it into a number of underlying related dimensions;

c) extending the dimensions into bipolar constructs by formulating opposites;

d) exemplifying the bipolar constructs by drafting pairs of indicative items; and

e) expressing each item pair in 'past' and 'present' forms.

The number of items in a scale was therefore determined by the number of underlying dimensions which could be distinguished. Items were drafted as statements describing experiences. Initially, it had been thought that these could best be rated on a 'true/not true' or 'true/false' scale. As it is, however, conceptually problematic to conceive of gradations of truth (for instance to evaluate a statement as 'somewhat true' or 'very true'), preference was given to a scale gauging the degree of accuracy of a descriptive statement, which lends itself more readily to a fine grained evaluation of an experience. The format eventually chosen consisted of the prompt: "To what degree do the following statements concur with your experience of the difficulty you have described?" rated on a seven-point-scale anchored by verbal descriptors ranging from "not at all" to "completely" (see 3.2.1.2.4 below). Care was taken in the formulation of items to exclude 'emotional qualifiers', such as *intimately* acquainted or *deeply* affected.

Such weighting adverbs would have required respondents in effect to make two separate rating decisions; one about the degree of congruence between a statement and their experience, and another one about the degree to which the quality denoted by the adverb was represented in their experience.

3.2.1.1.1.1 Familiarity

Dictionary definitions of familiarity usually refer to it as a state of being well known or of being closely acquainted. In the current context, familiarity was explicitly understood as a feeling of 'déja vu', a sense of having come across a difficulty previously, rather than as an experience of encountering it frequently in the present. The latter concept was taken up in a pair of items which were included in the final version of the instrument as 'additional items' outside the Familiarity and Impact scales (see 3.2.1.1.2 below) but which were thought to be inversely related to the paradigmatic difficulty concept[9].

In order to capture the full meaning of the concept, a set of eight associated dimensions - thought to represent close connotations of the main construct - were identified with the help of Roget's Thesaurus (Browning, 1972). These were labelled as 'Novelty', 'Frequency', 'Recency', 'Surprise', 'Uniqueness', 'Anticipation', 'Acquaintance', and 'Typicality'. For each of these dimensions a pair of items was drafted denoting its opposite poles. Positive polarity of an item connotes familiarity, negative polarity unfamiliarity. Individual items consisted of statements about experiences of difficulties which therapists might identify with. For reasons discussed below (see 3.2.1.2.1), there were two versions of each item, one in present tense and one in past tense.

---

[9]An idea that these items might be developed into a separate hypothesis, such as "paradigmatic difficulties are unlikely to be experienced with more than one patient at a time as therapists will learn to protect themselves against this kind of experience by selecting out clients which are likely to evoke it" was in the event not followed up.

The list of all 16 Familiarity items is displayed in Table 3:1 below.

Dimension	Polarity	Tense	Item
Novelty	+	present	This kind of difficulty is well-known to me.
		past	This kind of difficulty was well-known to me.
	-	present	Difficulties of this nature are new to me.
		past	Difficulties of this nature were new to me.
Frequency	+	present	I have experienced this sort of difficulty time and again.
		past	I had experienced this sort of difficulty time and again.
	-	present	This is virtually the first time that I have encountered a difficulty of this type.
		past	This was virtually the first time that I had encountered a difficulty of this type.
Recency	+	present	I have been acquainted with this sort of difficulty for a long time.
		past	I had been acquainted with this sort of difficulty for a long time.
	-	present	It is only recently that I first met this kind of difficulty.
		past	It was then only recently that I had first met this kind of difficulty.
Surprise	+	present	I did foresee (or could have foreseen) that I would encounter this difficulty.
		past	I did foresee (or could have foreseen) that I would encounter this difficulty.
	-	present	I am surprised to be confronted with this kind of difficulty.
		past	I was surprised to be confronted with this kind of difficulty.
Uniqueness	+	present	I am used to this kind of difficulty.
		past	I was used to this kind of difficulty.
	-	present	This is an unusual difficulty for me.
		past	This was an unusual difficulty for me.

Anticipation	+	present	I expect to meet this difficulty again in the future.
		past	I expected to meet this difficulty again in the future.
	-	present	I do not anticipate being faced with this kind of difficulty again.
		past	I did not anticipate being faced with this kind of difficulty again.
Acquaintance	+	present	I know about difficulties of this kind through previous experience.
		past	I knew about difficulties of this kind through previous experience.
	-	present	I am not personally conversant with this type of difficulty.
		past	I was not personally conversant with this type of difficulty.
Typicality	+	present	This is a typical difficulty for me.
		past	This was a typical difficulty for me.
	-	present	This difficulty is uncharacteristic of me.
		past	This difficulty was uncharacteristic of me.

Table 3:1 - Familiarity Scale: List of Items grouped by Dimension, Polarity and Tense

3.2.1.1.1.2 Depth of Impact

The second of the two new scales concerned the dimension of 'Depth of Impact' of a difficulty. Dictionary definitions of impact tend to focus on the idea of the collision of two moving bodies. In the current context, Depth-of-Impact of a difficulty was understood as denoting the degree to which it was capable of affecting the therapist experiencing it. This notion was meant to focus primarily on the strength of emotional experiencing in reaction to a difficulty, but initially it also encompassed aspects of impact on thought and imagery. When the scale was subsequently shortened - as described below in 5.2.1.1.3.3 - by retaining only those items which showed the

strongest internal consistency, these other perspectives dropped away, leaving a more homogenous 'Depth of Emotional Impact' scale.

There are some other impact scales reported in the literature. The IMI (Kiesler, Schmidt, & Wagner, 1997), used by Torres (1983) in her study (see 2.1.3 above) is a global measure of the interactional 'pull' of an interpersonal transaction rather than of the intrapsychic state of the participants. The 'Impact of Event Scale' (Horowitz, Wilner, & Alvarez, 1979), is a self-report measure tapping the response to stressful events along the two dimensions of avoidance and intrusive thinking. It has subsequently been developed into the 'Stress Response Rating Scale' (Weiss, Horowitz, & Wilner, 1984) a clinical observation measure, which elaborates the avoidance dimension into avoidance / denial and adds a general stress response dimension. However, both these measures focus on the psychological sequelae of stressful events rather than on the quality of experience during the event which is the point of interest in this study.

In order to capture the full meaning of the impact concept, a set of twelve associated dimensions - thought to represent close connotations of the main construct - were identified, again with the help of Roget's Thesaurus (Browning, 1972). These were labelled as 'Impact', 'Impression', 'Immersion', 'Contemplation', 'Preoccupation', 'Control', 'Penetration', 'Effort', 'Mindfulness', 'Dissociation', 'Saturation', and 'Concern'. As before, each of these dimensions was represented by a pair of items denoting its opposite poles; positive polarity indicating high impact, negative polarity low impact. Individual items consisted of statements about experiences of difficulties with which therapists might identify. Once more, there were two versions of each item, one in present tense and one in past tense.

The list of all 24 Depth-of-Impact items is displayed in Table 3:2 below.

Dimension	Polarity	Tense	Item
Impact	+	present	I am emotionally affected by this difficulty.
		past	I was emotionally affected by this difficulty.
	-	present	I am untouched by experiencing this difficulty.
		past	I was untouched by experiencing this difficulty.
Impression	+	present	I experience this difficulty as 'going to the core' of me.
		past	I experienced this difficulty as 'going to the core' of me.
	-	present	I experience this difficulty as going only 'skin-deep'.
		past	I experienced this difficulty as going only 'skin-deep'.
Immersion	+	present	I am engrossed by this particular difficulty.
		past	I was engrossed by this particular difficulty.
	-	present	I can ignore this difficulty.
		past	I could ignore this difficulty.
Contemplation	+	present	This difficulty gives me pause for thought.
		past	This difficulty gave me pause for thought.
	-	present	I do not feel the need to reflect on this difficulty.
		past	I did not feel the need to reflect on this difficulty
Preoccupation	+	present	I have been preoccupied by the experience of this difficulty outside the therapy session(s).
		past	I was preoccupied by the experience of this difficulty outside the therapy session(s).
	-	present	I am absorbed in this difficulty only when I am with my client/patient.
		past	I was absorbed in this difficulty only when I was with my client/patient.
Control	+	present	Thoughts, feelings or images connected with this difficulty come to mind unbidden.
		past	Thoughts, feelings or images connected with this difficulty came to mind unbidden.

	-	present	I can easily push aside the thoughts and feelings aroused in me by this difficulty.
		past	I could easily push aside the thoughts and feelings aroused in me by this difficulty.
Penetration	+	present	This difficulty has really 'got inside' me.
		past	This difficulty really 'got inside' me.
	-	present	I am able to distance myself from this particular difficulty.
		past	I was able to distance myself from this particular difficulty.
Effort	+	present	Grappling with this difficulty is a struggle.
		past	Grappling with this difficulty was a struggle.
	-	present	I find this difficulty easy to deal with.
		past	I found this difficulty easy to deal with.
Mindfulness	+	present	I feel concerned about this difficulty.
		past	I felt concerned about this difficulty.
	-	present	I feel untroubled by this difficulty.
		past	I felt untroubled by this difficulty.
Dissociation	+	present	I cannot let go of this difficulty.
		past	I could not let go of this difficulty.
	-	present	It is easy to disengage from this difficulty.
		past	It was easy to disengage from this difficulty.
Saturation	+	present	I am full of the experience of this difficulty.
		past	I was full of the experience of this difficulty.
	-	present	This difficulty stays with me only fleetingly.
		past	This difficulty stayed with me only fleetingly.
Concern	+	present	I am worried about this difficulty
		past	I was worried about this difficulty
	-	present	This difficulty does not bother me
		past	This difficulty did not bother me

Table 3:2 - Impact Scale: List of Items grouped by Dimension, Polarity and Tense

3.2.1.1.2 Construction of a Combined Scale

As the Familiarity and Impact scales were thought to be conceptually related (by positioning difficulties on the transient/paradigmatic spectrum), both sets of items were combined in a single measure for inclusion in TDQ1. Two items representing the idea of a difficulty occurring presently with more than one client (see 3.2.1.1.1.1 above) were also included and are shown in Table 3:3 below.

Polarity:	Tense:	Item:
+	present	My other existing patients/clients are not difficult in this particular way.
+	past	My other existing patients/clients are not difficult in this particular way
-	present	Currently, several of my patients/clients present me with this difficulty
-	past	At the time, several of my patients/clients present me with this difficulty

Table 3:3 - Additional Items included in the Combined Scale

In order to avoid order effects, items were positioned in the combined scale so that positive and negative versions of each item were placed approximately equidistant from the middle of the scale and that negative polarity was presented first for half the items and second for the other.

Table 3:4 below shows the order of items in the combined scale, indicating their polarities and the concepts and dimensions they represent. Items versions relate to the present-tense form. The past-tense form is exactly equivalent.

Polarity:	Scale:	Dimension:	Item:
+	IMP	Impact	I am emotionally affected by this...
-	FAM	Novelty	Difficulties of this nature are new..
+	IMP	Impression	I experience this difficulty as...
-	Additional		My other existing patients/clients..
-	IMP	Immersion	I can ignore this difficulty.
+	FAM	Frequency	I have experienced this sort of...
+	IMP	Contemplation	This difficulty gives me pause...
-	IMP	Preoccupation	I am absorbed in this difficulty...
+	FAM	Recency	I have been acquainted with this...
+	IMP	Control	Thoughts, feelings or images...
-	FAM	Surprise	I am surprised to be confronted...
+	IMP	Penetration	This difficulty has really 'got...
-	FAM	Uniqueness	This is an unusual difficulty for...
-	IMP	Effort	I find this difficulty easy to deal...
-	IMP	Mindfulness	I feel untroubled by this difficulty.
+	FAM	Anticipation	I expect to meet this difficulty...
-	IMP	Dissociation	It is easy to disengage from this...
+	FAM	Acquaintance	I know about difficulties of this...
+	IMP	Saturation	I am full of the experience of this..
-	FAM	Typicality	This difficulty is uncharacteristic...
+	IMP	Concern	I am worried about this difficulty.
+	IMP	Effort	Grappling with this difficulty is a...
+	FAM	Typicality	This is a typical difficulty for me.
-	IMP	Saturation	This difficulty stays with me only..
-	FAM	Acquaintance	I am not personally conversant...
+	IMP	Dissociation	I cannot let go of this difficulty.
-	FAM	Anticipation	I do not anticipate being faced...
+	IMP	Mindfulness	I feel concerned about this...
-	IMP	Concern	This difficulty does not bother me.
+	FAM	Uniqueness	I am used to this kind of difficulty.

-	IMP	Penetration	I am able to distance myself from..
+	FAM	Surprise	I did foresee (or could have...
-	IMP	Control	I can easily push aside the...
-	FAM	Recency	It is only recently that I first met...
+	IMP	Preoccupation	I have been preoccupied by this...
-	IMP	Contemplation	I do not feel the need to reflect...
-	FAM	Frequency	This is virtually the first time that..
+	IMP	Immersion	I am engrossed by this difficulty.
+	Additional		Currently, several of my patients...
-	IMP	Impression	I experience this difficulty as...
+	FAM	Novelty	This kind of difficulty is well-...
-	IMP	Impact	I am untouched by experiencing...

Table 3:4 - Ordering and Polarities of Items in Combined Scale

Ratings for items were conceived as forming a seven-point ratio scale ranging from '0' to '6'. Scale points were anchored verbally in the following way: 0 - 'not at all', 1 - 'marginally', 2 - 'somewhat', 3 - 'partially', 4 - 'substantially', 5 - 'almost entirely', 6 - 'completely'. The ordering of these descriptors had been confirmed by having three native English-speakers rank-order them which resulted in complete agreement. The ratio quality of the scale (Everitt, 1996), which is not implied in the ordering of verbal anchors, was conveyed in three ways: a) by the numbering which implies equidistant scale points and an absolute zero; b) by presenting numerical points for each item in an array with equal distances between numbers, resembling a visual analogue scale; and c) by visually representing scale points with pie-charts depicting increasing coverage of a circle (see appendix 2).

3.2.1.1.3 The Duration Index

Duration, defined by the Oxford Concise Dictionary as "the length of time for which something continues", was in the present context understood as a retrospective estimate about how long a difficulty had persisted. In the case of 'current or very recent' difficulties this might be the interval between the first occurrence of the difficulty and the point at which the respondent completed the questionnaire. For such ongoing difficulties, the respondent's estimate about the likely future duration of the difficulty was added to the past duration.

Although a number of other items were included in the relevant section (see 3.2.1.2.4 below), the Duration Index was in the end simply calculated in weeks from the answers to the following items (see appendix 2):

P 3.2. "For how many weeks did the difficulty last altogether?" (past difficulties), or

C 3.6. "How many weeks did it last altogether?" (current concluded difficulties), or

C 3.2. "For how many more sessions[10] do you expect it to persist or recur?" plus

C 3.4. "How many weeks ago did it first occur?" (ongoing current difficulties).

An alternative to this 'objective' index would be a scale consisting of items tapping therapists' subjective experience of duration. This possibility was considered but, in the event only followed up in the design of TDQ2 (see 5.2.1.1.3.1 below).

[10] This logical break (sessions rather than weeks) is a design fault in TDQ1. It is mitigated by the fact that most reported difficulties occurred in weekly therapy.

3.2.1.2 Questionnaire Construction

3.2.1.2.1 Rationale

The investigations with TDQ1 rested on the assumption, that therapists can call difficult situations to mind when prompted to do so, and will be able to answer retrospectively questions relating to their experiences of the difficulties. Asking respondents to provide written accounts of difficult situations (in the way already trialled in the survey of British psychotherapists reported in 1.2 above) had originally been envisaged as a method of cueing respondents into their past experiences, facilitating their detailed recall of situations, and helping them to answer related questions. Subsequently, the narratives themselves moved into focus; initially with a view to formal content analysis, but later as a basis for the ratings in the second study, described below in 4.2.4.

Asking for the recall of difficult situations in two different formats - a 'past difficulty' and a 'current or very recent difficulty' - was meant to enhance the richness of the database by giving respondents two distinct contexts as cues for their recall. An additional reason was the thought that 'past difficulties' might more likely be located towards the paradigmatic end of the proposed spectrum because they would have to have had considerable emotional impact at the time if they could be recalled in the present.

3.2.1.2.2 The Cover Page

The cover page of TDQ1, which was varied for different samples (see appendix 2) provided an opportunity to engage potential respondents in a research alliance. (In postal samples, this was supplemented by a covering letter, see appendix 3.) It

thanked potential respondents for considering participating in the study, outlined potential benefits to the investigator and to respondents, drew attention to the optional part of the questionnaire, gave information as to how it could be returned, and explicitly invited feedback.

3.2.1.2.3 Eliciting Accounts

Accounts for both current and past difficulties were elicited in an equivalent format, each on a single A4 page. First respondents were asked to think of a situation (either "current or very recent" or having occurred "some considerable time ago") which they had personally encountered in their practice of individual psychotherapy and which they had found difficult. They were then provided with spaces to write down their accounts under the same headings previously used in the survey of British psychotherapists reported in 1.2 above:

1. What did you or your patient (client) do which made the situation difficult?

2. What feelings or personal reactions did you experience in the situation?

3. How did you attempt to deal or cope with this difficulty?

4. How did this situation turn out?

These sections of the questionnaire were displayed on single pages to enhance clarity of presentation and to save space. A possible disadvantage of such a design lies in the limited space afforded to the answers to each subheading which might encourage briefer replies containing less of the relevant information. To counteract such a restriction respondents were invited to also use the reverse of the page to continue their account if necessary.

Half the questionnaires distributed started with the 'past difficulty' account, the other half with the 'current or very recent difficulty', to guard against possible order effects.

3.2.1.2.4 FID Scales

The measures for Familiarity, Impact, and Duration (FID Scales), in their appropriate versions, were placed immediately after the difficulty accounts under the heading "ABOUT YOUR EXPERIENCE OF THE DIFFICULTY". Initially, the list of 42 items had only been preceded by the question "To what degree do the following statements concur with your experience of the difficulty you have described?", supplemented in the 'current' version with "In the case of a very recent difficulty give the response as you felt at the time.", and followed in both versions with the instruction "(Please circle one number for each statement)". Verbal anchors for scale points were provided in a separate box.

Because of feedback from early respondents that this format was not sufficiently clear, later versions of the questionnaire also contained at the top of the list of items the rider "This statement captures my experience...", which was visually linked to the list of verbal descriptors ranging from 'not at all' to 'completely'.

Section 3 contained the items used for the calculation of the Duration Index under the heading "ABOUT THE DIFFICULTY", whereas section 4 - under the heading "ABOUT THE THERAPY IN WHICH THE DIFFICULTY OCCURRED" contained some further items relating to frequency of sessions and overall treatment length. These could have been used in the calculation of a more complicated Duration index but were, in the event, discarded from further analyses.

3.2.1.2.5 Patient Variables

Section 5 of TDQ1 - items 5.1 to 5.3 - elicits information about age, sex, and level of impairment for each of the two patients with whom the difficulties occurred. The first two items were included to provide minimal descriptive information about the patient

sample, but also with the idea in mind that therapist-patient matches on these variables might be related to the transient-paradigmatic spectrum. The third item, measuring levels of patient disturbance as 'mild', 'moderate' or 'severe', was included because of the possibility that it might differentiate between transient and paradigmatic difficulties .

3.2.1.2.6 Therapist Variables

Five therapist variables - items 6.1 to 6.5 - were included in section 6 of the questionnaire - under the heading "ABOUT YOURSELF" - with the aim of providing a basic description of the therapist sample. Apart from age and sex, information about length of practical experience, profession, and theoretical orientation was elicited. Allowing for more than one choice, the six professions of which respondents were most likely to be members (psychologist, psychiatrist, psychotherapist, social worker, counsellor, and nurse) were presented as a list together with an open-ended 'other' category. Five (not mutually exclusive) choices of theoretical orientations (psychodynamic, behavioural, cognitive, humanistic, and systemic) were presented in a similar fashion, again supplemented by an open category. The item enquiring into length of practical experience was presented as "For how many years have you been practising as a therapist?" together with an answering template "approx. . . . . years". Further possible qualifiers, such as "since qualification" or "since finishing your basic training", were deliberately omitted from the question as the variety of ways in which individuals can become therapists would have made it impossible to define a generalisable starting point for the accumulation of therapeutic experience. The wording of the item therefore leaves the definition of practice length to respondents themselves, taking as a common yardstick therapists' felt sense of the duration of their practice.

3.2.1.2.7 Difficulties Scale and Coping Strategies Scale

The Difficulties Scale, which forms section 7 of the questionnaire, is identical to the measure used in the CCQ of the ISDPD, described in 2.2.1.2.2 above. The main reasons for its inclusion were unconnected with the current study and it was therefore designated as 'optional'. In the event, it did yield some interesting comparative information about the level of difficulties reported by TDQ1 respondents (see 3.3.5 below) and it provided a potential control or mediating variable for use in further data analyses.

Likewise, the Coping Strategies Scale, forming section 8 of the questionnaire, is the same as that used in the CCQ. It too, was peripheral to the questions under investigation in this study and was therefore also incorporated into the 'optional' section of TDQ1. Its only use for the current study is as a potential control or mediating variable.

3.2.1.2.8 Feedback Section

The feedback section consisted of two components: a structured part with three multiple choice items and an open-ended part inviting written comments on the questionnaire and the respondent's experiences while completing it.

3.2.1.3 German Translation

A translation of the entire questionnaire into the German language was prepared with the intention to collect data from German speaking therapists in order to

a) obtain a broader sample (if both versions of the questionnaire were to yield similar results in terms of the psychometric properties of the scales) in case of there not being a large enough response to the English language survey;

b) provide a convenient way of replicating results found in one of the language-samples - such as intercorrelations or associations of the scale with patient or therapist variables - in the other language-sample.

c) supply pointers to possible cultural differences if the psychometric properties of the scales were to be dissimilar or if associations etc. were not to be replicated.

The translation was carried out by myself as German is my first language. A draft version was checked by two native German-speakers for linguistic correctness and by a bilingual Clinical Psychology trainee for accuracy of translation. The final questionnaire version was piloted with a native German-speaking psychotherapy researcher in order to ascertain that technical terms had been translated correctly and to assess the general acceptability of the questionnaire to a German-speaking respondent.

An attempt to collect data from a German speaking sample by distributing approximately 30 of the translated questionnaires among the audience of a paper on therapist difficulties given at the 1993 European SPR meeting in Budapest (Schröder, 1993) was unsuccessful as none were returned. The reasons for this complete lack of response are difficult to ascertain, but one might speculate that they can partly be found in the nature of the audience whose main preoccupation in this first SPR conference in a former Eastern bloc country may have been to gather and exchange information rather than to participate in current research. Another reason may lie in the fact that the conference language was English, making it less likely for potential German respondents to be directly cued into the questionnaire they were provided with.

As the number of English language responses had at the time begun to approach an acceptable level, no further attempts at data collection with the German TDQ1 were

made. However, the effort expended in preparing the translation were not entirely wasted, as substantial portions of the questionnaire were used in the German version of TDQ2 and as the experience gathered with the former assisted in the preparation of this latter questionnaire. The experience of the failed data collection at the Budapest conference also prompted a rethink of the sampling strategy for TDQ2.

3.2.2 Data Collection

Pilots of TDQ1 had shown it to be quite lengthy to complete - half an hour appearing a reasonable estimate, although extended difficulty narratives could considerably prolong the time required, leading to lowered expectations of return rates. The main concern in collecting data was therefore to recruit enough subjects into the study to allow analyses of the psychometric properties of the scales and of the associations between Impact, Familiarity and Duration and other variables. Given the experiences from the Warwick Collaborative Study, a target of 100 respondents was thought to be the maximum to be hoped for; however, in view of the length of TDQ1, a lower target seemed more realistic. In the end a total of 50 completed questionnaires was adopted as the target to be reached before further efforts at data collection would cease.

3.2.2.1 Sampling

No attempts were made to obtain a 'representative sample' of psychotherapists, partly for conceptual reasons, as the population from which samples could be drawn is of unclear delineation and unknown characteristics, and partly for practical reasons, as the expected low return rate renders all generalisations problematic (as discussed below). Instead, the strategy adopted was one of seeking out a group which could be expected to be responsive to an appeal for co-operation, of trying to maximise returns

by finding additional means of engaging respondents, and of selecting further groups until the target of fifty subjects in the study was accomplished.

3.2.2.1.1 Conference Sample

The first data collection for this enquiry was undertaken at the 10th Annual Meeting of the British chapter of the Society for Psychotherapy Research (SPR(UK)). A paper (Schröder & Davis, 1993), outlining the conceptual background of the study but keeping the audience unaware of the specific hypotheses then under investigation, had been scheduled for the first day of the conference. Following the presentation and subsequent discussion, those present were invited to take a copy of TDQ1 and return it either during the conference (anonymously using a collection box in the main conference room) or later on by post. The entire first print run of 60 questionnaires was distributed on this occasion. Only 4 questionnaires were returned during the conference, a further 10 arriving later by mail.

3.2.2.1.2 Previous Survey Respondents

A second target group for data collection consisted of previous respondents to the survey of British Psychotherapists described in section 1.2 above. Out of a total of 96 respondents, 80 had supplied their names and addresses by returning a reply form separately from the main questionnaire; 72 of these were contacted, the remaining eight were thought to be unavailable. All of the contacts received two copies of TDQ1 together with a covering letter (see appendix 3) making reference to the benefits of their previous contributions and inviting them to pass on the spare copy to a colleague who might be interested in participating. The accompanying response form (see appendix 3), providing an estimate of the return rate in this sample, was sent back by 22 of the subjects contacted in this way. Nine of the previous respondents wrote to

indicate that they were unable to participate. Reasons given included retirement and changes in work patterns.

3.2.2.1.3 Convenient Local Sample

The third group targeted to provide data for this stage of the research project consisted of local professional colleagues and fellow psychotherapists who could be expected to be accommodating or to feel a sense of personal obligation. In addition, two smaller groups were canvassed collectively: the first comprising members of the Clinical Psychology Department of one of the local trusts, who filled in the questionnaire in the course of a workshop on therapist difficulties; the second consisting of visiting therapists in a neighbouring psychotherapy unit, who were invited to take away copies of TDQ1 following a lecture in the course of a guest speaker event. Approximately 60 questionnaires were distributed to this sample.

3.2.2.2 Return Rates

Return Rates for the three samples are summarised in Table 3:5 below.

Sample	Questionnaires distributed	Questionrs. / Forms returned	Return Rate
Conference	**60**	**14**	**23%**
Previous Respondents	**72**	**22**	**31%**
Local Sample	**60**	**24**	**40%**

Table 3:5 - TDQ1: Return Rates

Not surprisingly, these are highest for the local sample, where there had been the best opportunity of building a research alliance, and lowest for the conference sample. Even the best return rate is quite modest, indicating the problems of collecting data with a relatively lengthy instrument. This experience led to strenuous efforts to engage potential respondents in the subsequent study reported in chapter five below.

3.2.2.3 Ethical Considerations

Ethical issues raised by the data collection with TDQ1 are very similar to those surrounding the subsequent questionnaire, TDQ2. They are therefore discussed together in 5.2.2.2 below. Like its successor, TDQ1 was not submitted for approval to an Ethical Committee.

3.2.3 Data Entry

3.2.3.1 Format and Manual

Numerical data were entered into a single computer work file by using the data editor of the SYSTAT statistical software package. Part of the data-entry was carried out by an assistant following a manual (see appendix 4) which specifies variable names and datapoint numbers and gives directions how to input data for each variable. Marks or ticks falling between scale points were assigned the lower of the two values. Other ambiguous responses were discussed with the assistant and those which remained unclear were entered as missing values.

The written accounts and any written comments in response to item 9.4. were entered into separate textfiles using the DR DOS text editor. This generated a maximum of three files per respondent (past difficulty account, current difficulty account and comments). Printouts of the textfiles were proof-read against the originals, discrepancies in the interpretation of words which were hard to decipher were resolved in discussions (occasionally enlisting the help of colleagues when a respondent's handwriting was particularly intractable), and words which proved totally illegible were entered as one '?' per word. Idiosyncrasies in spelling were retained as written in the originals as were abbreviations such as '+' or '&' in place of

'and'. Paragraphs appearing in the original were retained; however, indents and unusual positionings of text were not preserved at this stage.

3.2.3.2 Verification of Datafile

In order to check that numeric data had been entered correctly from the questionnaires into the computer file, the following procedures were adopted:

a) Summary statistics and density plots for all variables were printed out and checked for illegal values (those lying outside the range of possible values for a given variable) and implausible values. Those found were traced in the questionnaires to ascertain whether they were invalid responses or data-entry errors. No invalid responses and seven data-entry errors were detected, the latter representing less than 0.1 % of all 10431 numeric datapoints.

b) Six questionnaires were selected at random and compared against the respective listings from the datafile. One data-entry error was trapped in this way, representing less than 0.1% of the 1098 datapoints thus examined.

c) All missing data entries and all 'inconsistent responses' (for definition see 3.3.4.1.1.) for the Familiarity and Impact scales from the datafile were compared against the original questionnaire entries. One data-entry error was discovered, representing approximately 0.3% of the 316 datapoints checked in this way; however, an investigation as to why this mistake should have arisen led to the detection of a further 11 data-entry errors in the same record, which had occurred because a section of the data had been right-shifted by one column in the fixed-format computerfile. As sample items of most questionnaires were verified, it seems very unlikely that another data-entry error of this type would have remained unnoticed.

As a result of these checks, there is reason to have some confidence in the accuracy of the computer data file as a true representation of the raw numeric questionnaire data.

3.2.4 Data Analysis

3.2.4.1 Consistency of Responses

A few of the early respondents to TDQ1 had reported finding some double negatives in the 'familiarity' and 'impact' scales confusing; giving rise to the concern that both items in a pair might erroneously have been answered in the same direction. As items were designed to denote opposite poles of constructs and rated on a 0 - 6 scale, 'consistent' responses to each item-pair representing a construct (either in present or past form) should ideally sum to 6.

The following definitions were adopted to discriminate between responses to item-pairs:

'Consistent' responses were defined as those which yielded sums in the range of 4 to 8;

'Inconsistent' responses were defined as those which added up to 0, 1, 2, 10, 11, or 12;

'Marginally consistent' responses were defined as those which summed to 3 or 9.

Sums of all responses to all item-pairs were computed and assigned to one of the above categories in order to investigate whether certain respondents had raised rates of inconsistent responses, suggesting that they were more prone to misinterpreting the scaling instructions (as some of the written comments on the questionnaires had suggested).

The pattern of individuals' responses to the questionnaire items was explored first, in order to exclude those with high numbers of inconsistent ratings before item analysis. Summary statistics for inconsistent responses are shown in the first column, those for

inconsistent and marginally consistent responses are shown in the second column of Table 3:6 below.

N (Respondents) = **57** **N** (Responses) = **84**	Inconsistent Responses	Inconsistent and Marginally Consistent Responses
MEAN	**4.4**	**8.5**
SD	*3.2*	*4.3*
Min	0	1
Max	14	19
MEDIAN	***4***	***8***

Table 3:6 - Inconsistent Responses: Descriptive Statistics

As there were 42 item pairs altogether (21 each for past and present versions), the mean for inconsistent responses represents just over 10% of all responses. If marginally consistent responses are included as well, the mean rises to just under 20%. The least consistent responder had one third of 'inconsistent' responses.

The cut-off points for excluding respondents was defined as two standard deviations above the mean on either variable. This would be **either** 11 (≈ 25%) or more inconsistent responses, **or** 18 (≈ 42%) or more combined inconsistent and marginally consistent responses.

Four respondents (nos. 22, 28, 33, and 38) were excluded on the basis of these criteria. A further respondent who came close to the cut-off point was excluded on the basis that more than half the scale ratings and substantial segments of other data were missing. The selection of items for the Familiarity and Impact scales is therefore based on N = 52 of the returned questionnaires. Summary statistics for inconsistent responses after the exclusion of the five most inconsistent respondents are shown in Table 3:7 below.

N (Respondents) = **52** **N** (Responses) = **84**	Inconsistent Responses	Inconsistent and Marginally Consistent Responses
MEAN	**3.8**	**7.7**
SD	*2.5*	*3.4*
Min	0	1
Max	9	15
MEDIAN	***3.5***	***7.5***

Table 3:7 - Inconsistent Responses: Descriptive Statistics after Exclusions

As one would expect, there is less variability shown by either measure of consistency.

Means and medians are reduced and the least consistent respondent is now someone with less than a quarter of all responses classed as 'inconsistent'.

In order to explore, whether the modification of the questionnaire (by making the scaling instructions more explicit) had an impact on the consistency of responses, T-tests were computed comparing the means for inconsistent and marginally consistent responses for both versions. Although both means were somewhat lower for the modified version, this difference did not reach statistical significance.

3.2.4.2 Other Issues

After checking for inconsistent responses, items with reverse polarity were recoded for both scales, so that for all items higher scores would be indicative of higher familiarity or impact ratings. In order to check the accuracy of the command file used for the transformation, two dummy questionnaires with extreme values in either direction were recoded and the results verified by inspection.

Missing data amounted to a total of 53 (1.2% of all Familiarity and Impact scale data points). They were not recoded but carried forward into subsequent analyses leading to a slightly reduced N.

In addition to the 57 questionnaires included in the consistency analysis, a further three questionnaires had been received, completed (in English) by respondents whose first language was not English (two Swedish and one Swiss). As it could not be assumed that the scale items would have been understood by them as they would be by native speakers, data from foreign-language respondents were excluded from the study.

3.3 Results

3.3.1 Sample Characteristics

The sample of respondents used for data analyses can be described in terms of the variables contained in section 6 of the questionnaire. Responses to these items are summarised in Table 3:8 below. The sample is balanced between male and female respondents. Statistics for age and length of practical experience show good range and variability and demonstrate that the sample was mature both in terms of life experience and of professional experience. Percentages for profession and for theoretical orientation sum to more than 100 because of multiple endorsements. Psychologists and therapists with a psychodynamic orientation each account for about three quarters of the total sample. Thus, the typical respondent to TDQ1 was a psychodynamically orientated psychologist, about 40 years of age, with over ten years of clinical experience; in short, someone much like myself, which is not surprising given that this is a 'convenient' sample.

TDQ1 Respondents (N=52)		
Sex:		**Female: 51%** **Male: 49%**
Age:	Mean (SD):	**39.3** (7.6)
	Median:	**40**
	Range:	**26 - 63**
Practice Length (Years):		
	Mean (SD):	**11.0** (6.3)
	Median:	**10**
	Range:	**1 - 30**
Profession		
	Psychologist:	**71%**
	Psychotherapist:	**24%**
	Psychiatrist:	**16%**
	Counsellor:	**6%**
	Social Worker:	**6%**
Theoretical Orientation		
	Psychodynamic:	**73%**
	Cognitive:	**39%**
	Humanistic:	**20%**
	Behavioural:	**16%**
	Systemic:	**16%**

Table 3:8 - TDQ1 Respondents: Sample Characteristics

3.3.2 Item Analyses and Scale Reliabilities

The main goal of item analysis in this study was to modify scales in order to maximise reliabilities. Shortening of scales and eliminating items showing poor discrimination (indicated by very high or very low means and skewed distributions) was not of primary interest here, but became so in a later study (see 5.2.1.1.3.2 and 5.2.1.1.3.3 below). Data files were imported into the SPSS statistical software package and analysed in the 'Scales' module, using Cronbach's alpha as the measure

of internal consistency. A striking feature of all scales were the high alphas of the total, unmodified scales.

3.3.2.1 Past Familiarity

The initial internal consistency for the unmodified scale was estimated as $\alpha = .90$. Means for all but two individual items ranged from **1.56** (*I had experienced this sort of difficulty time and again*) to **3.5** (*I was surprised to be confronted with this kind of difficulty* - negative polarity). The remaining two items, both representing the 'Anticipation' dimension had markedly higher means - **3.9** (*I expected to meet this difficulty again in the future*) and **4.7** (*I did not anticipate being faced with this kind of difficulty again* - negative polarity). These two items also depressed the overall scale reliability and were eliminated from further analysis, leaving a 14-item modified Past Familiarity scale with an estimated internal consistency of $\alpha = .92$.

3.3.2.2 Current Familiarity

The initial internal consistency for the unmodified scale was estimated as $\alpha = .96$, indicating that familiarity is somewhat more difficult to measure reliably if a difficulty occurred "some considerable time ago". Means for all but the two 'Anticipation' items ranged from **1.96** (*I have experienced this sort of difficulty time and again*) to **3.75** (*I was surprised to be confronted with this kind of difficulty* - negative polarity). Again, the 'Anticipation items had markedly higher means - **4.3** and **4.6** - and marginally depressed overall scale reliability. Eliminating them from further analysis left a 14-item modified Current Familiarity scale with an estimated internal consistency of $\alpha = .97$.

3.3.2.3 Past Impact

The initial internal consistency for the unmodified scale was estimated as $\alpha = .93$. Means for individual items ranged from **2.94** (*I was full of the experience of this difficulty*) to **5.21** (*I felt untroubled by this difficulty* - negative polarity). Scale reliability could not be improved by discarding any items and the Past Impact scale was therefore retained unmodified.

3.3.2.4 Current Impact

The initial internal consistency for the unmodified scale was estimated as $\alpha = .95$. Means for individual items ranged from **1.79** (*I am full of the experience of this difficulty*) to **4.85** (*I do not feel the need to reflect on this difficulty* - negative polarity). All items contribute to internal consistency and the Current Impact scale was therefore retained unmodified.

3.3.2.5 Modified Scales used for further Analyses

The only modification made to the scales used for further data analyses, was the exclusion of the item pair representing the 'Anticipation' dimension from both versions of the familiarity scale. Internal consistencies are high, ranging from .92 to .97 (see Table 3:9 below), and suggesting that there might be considerable scope for shortening all four scales without jeopardising reliability. The additional items showed no interpretable associations to any of the scales and were discarded.

3.3.3 FID Scales - Descriptive Statistics

Means, Standard Deviations, and medians for the modified Familiarity Scales, Impact Scales and the Duration Index (expressed in weeks) together with scale reliabilities are shown in Table 3:9 below.

	α	Mean	SD	*Median*
Past Familiarity	**.92**	**2.34**	1.24	***2.4***
Current Familiarity	**.97**	**3.11**	1.60	***3.1***
Past Impact	**.93**	**4.24**	.88	***4.3***
Current Impact	**.95**	**3.28**	1.09	***3.1***
Past Duration		**22.1**	26.8	***10***
Current Duration		**19.7**	22.3	***12***

Table 3:9 - FID Scales: Descriptive Statistics

Inspection of probability plots showed that the Familiarity and Impact scales are approximately normally distributed around their respective means. Distributions for the Duration Indices are highly skewed (as indicated by the differences between medians and means) and log transformations were applied for all further analyses. The mean rating for Current Familiarity is significantly higher than that for Past Familiarity ($t = -3.12$; $df = 50$; $p = .003$), presumably because of the problems in rating familiarity some considerable time after the event. Conversely, the Past Impact mean is significantly higher than the Current one ($t = -5.08$; $df = 50$; $p < .0005$), most likely because higher impact difficulties are easier to recall at some distance. The difference between Past and Current Duration means is not significant.

3.3.4 Associations Amongst FID Measures

Intercorrelations between Familiarity, Impact, and Duration measures are shown in Table 3:10 below. Figures in *italics*, above the diagonal, relate to the 'current', figures in normal script, below the diagonal, relate to the 'past' versions. In view of the high internal consistencies for the scales no corrections for attenuation have been applied.

			CURRENT	
		Familiarity	**Impact**	**Duration**
PAST	**Familiarity**	1	***-.27***	***.17***
	Impact	**-.02**	1	***.18***
	Duration	**.12**	**.12**	1

Table 3:10 - FID Scales: Intercorrelations

There are slight positive associations in evidence between the Impact Scale and the Duration Index, in both 'past' and 'current' versions. Past Impact and Past Familiarity are effectively independent of each other, whereas the 'current' versions of these scales show a modest negative correlation - unfamiliar current difficulties appear to have higher impact on the therapists experiencing them. While this is not entirely counter-intuitive, it does not accord with the hypothesis that high scores on all three measures are associated with paradigmatic difficulties, which would have led one to expect moderate positive associations between all three measures.

A formal test of the hypothesis is possible through comparison of actual with expected co-occurrences of high and low FID scale values. If scores on each of the three measures are divided into three equal categories ('high', 'middle', and 'low'), then one would expect the probability of a score falling simultaneously into all three 'high' or all three 'low' categories to be one in twelve (3*2*2) each. With an N of 51 this would give an expected number of 4.25 difficulties falling into the combined 'high Impact'/ 'high Familiarity'/ 'high Duration' category and the same expected number for the combination of the three 'low' categories. The actual numbers are as follows:

'Past' high FID: **1**; 'past' low FID: **5**; 'current' high FID: **2**; 'current' low FID: **3**.

As these values are no better than chance, it is unlikely that a combination of low scores would indicate transient and a combination of high scores paradigmatic difficulties. The hypothesis outlined in 3.1 is therefore not supported.

3.3.5 Difficulty Scale - Comparisons with other Samples

As mentioned above, the Difficulty Scale is not of primary interest in this study. It did, however, yield one result which merits discussion in the current context. The mean score on the 20-item inventory, where therapists are rating the frequency with which they have experienced a particular difficulty on a '0' - '5', 'never' to 'very often' scale), can be taken as a measure of a respondent's general propensity to experience difficulties in the course of their work. Comparisons of the mean for TDQ1 responders with two other samples are shown in Table 3:11 below. The CCQ 1630 sample consists of all respondents to the Common Core Questionnaire of the ISPDP (see 2.2.1.2.2 above) who had returned data by the beginning of 1994. A subsample, labelled CCQ(UK), consists of all respondents resident in the UK and is similar in size to the TDQ1 sample.

	N	Mean	*SD*
TDQ1	**50**	**2.04**	*.51*
CCQ (UK)	**43**	**1.33**	*.50*
CCQ 1630	**1630**	**1.44**	*.60*

Table 3:11 - Difficulty Scale: Comparison with two other samples

The mean frequency with which respondents experienced difficulties is significantly higher in the TDQ1 sample than in the CCQ(UK) sample ($t = -5.69$; $df = 41$; $p < .0005$), raising the question whether participants in the current study had a higher propensity to experience difficulties. It is possible, especially in view of the low return rates, that there is a positive association between therapists' proneness to

experiencing difficulties and their willingness to return a relevant questionnaire. A more likely explanation would be, that TDQ1 respondents had been cued into thinking about their own difficulties by reporting in detail on two of them before completing the Difficulty Scale at the end of the questionnaire, and therefore had higher estimates of their frequencies. Such a bias could only have been avoided by placing the Difficulty Scale at the beginning of the questionnaire, which would have had the disadvantage of cueing respondents into particular types of difficulty descriptions before they answered the free-format questions.

3.4 Discussion

The study based on the survey with TDQ1 succeeded in some respects, but failed in others. On the positive side, highly reliable measures of two experiential associates of difficulties (familiarity and depth of impact) had been constructed and tested out in practice. Furthermore, a substantial number of accounts of difficulties had been collected and were available for further analyses. Finally, valuable experience about the problems involved in engaging respondents to fill in a lengthy questionnaire had been gathered. On the negative side, the planned cross-cultural comparison had not been possible, because the collection of German-language data had foundered. More importantly, the hypothesised co-occurrence of extreme levels of familiarity, impact, and duration had not been observed, leaving the original research plan in some disarray. Bearing the well-worn (but nevertheless accurate) therapeutic cliché that "there is no cure for reality" in mind, it became clear that a different strategy of investigation was called for. This is described in the next chapter.

4. Chapter Four: Survey 1 - The Second Study

This second study used the same material as the first - difficulty accounts and associated measures collected with TDQ1 - but employed quite a different strategy to analyse those data. It prompted a reformulation and extension of the dimensions of 'pervasiveness' postulated in research question 1 (see 1.3 above) and led to theoretical clarifications informing the design and analysis of TDQ2 - the subsequent questionnaire which formed the basis of the third study in this thesis.

4.1 Rationale

Following the first analysis of the data collected with TDQ1, it had become apparent that the strategy of first seeking to identify paradigmatic and transient difficulties by the levels of familiarity, impact and duration associated with them, and then validating that classification by confirming theoretical predictions about associations with other measures, had led to inconclusive results. An alternative strategy suggested itself which would make use of the written accounts supplied by respondents. Originally (see 3.2.1.2.2 above), these had been elicited in order to cue respondents into the situation they were recalling and with a view to formal, 'objective' content analysis (Russell & Stiles, 1979). In addition, these narratives could also form the basis of classification into qualitative categories by way of ratings, analogous to the strategy adopted in the taxonomy development study (see 1.2 above). If it was possible to arrive at reliable classifications of accounts into 'transient' and 'paradigmatic' categories, these could substitute for Familiarity/Impact/Duration as a basis for further analyses and furthermore, their relationship to those dimensions could be investigated empirically.

4.2 Method

In order to utilise therapists' accounts in the way outlined above, it was first necessary to develop category definitions which were mutually exclusive, not confounded with the independent variables of interest in the study, and clear enough to be used reliably by external raters. Following on from this, the question needed addressing whether the accounts should be presented 'context-free' or accompanied by additional information to help raters in their task. After deciding on the format in which the material was to be presented and designing instructions for the rating task, the question of whether to train raters had to be addressed prior to recruiting a pool of raters and presenting them with the task. Subsequently, reliabilities of ratings had to be established and, if these were found to be generally satisfactory, subgroups of raters had to be selected in order to maximise reliabilities for the assignment of mean category scores to difficulty accounts. The issue of the validity of the difficulty types as defined was then explored by comparing ratings with those of the previously developed difficulties taxonomy and examining the results for predicted associations. Finally, the relationship of the ratings to the measures of Familiarity, Impact and Duration could be established and their associations to other variables investigated.

4.2.1 Difficulty Definitions

As 'episodic' difficulties had previously been thought of as occupying a middle section on a 'transient' - 'paradigmatic' continuum, their conceptual basis was not distinctly defined, being neither one nor the other but potentially containing elements of both. It seemed therefore more promising to concentrate on clarifying the definitions for transient and paradigmatic difficulty types and to rate accounts in terms of these polarities. In the event, a third type of difficulty, lying outside the

proposed continuum suggested itself during this process (see 4.2.1.3 below). Care had to be taken to exclude any connotations with the familiarity, impact and duration concepts from the definitions, as these would have confounded empirical associations.

4.2.1.1 Transient Difficulties

The central defining feature of transient difficulties is their connection with deficits in the knowledge or technical repertoire of the therapist who is experiencing them. Depending on one's frame of reference such deficits may be understood, quite specifically, as a lack of skills; more broadly, as an insufficiency of theoretical knowledge or of clinical comprehension; or, most inclusively, as a shortage of accumulated therapeutic experience. Not only are such deficits present, they are highlighted by the situation described and are instrumental in that situation being experienced as difficult. If therapists were to broaden their therapeutic repertoires by acquiring further skills or knowledge or by accumulating more clinical experience, one would expect such transient difficulties to subside or disappear altogether until they encountered new situations for which their current repertoires were not adequate. As transient difficulties are closely associated to career status, they are likely to be generalisable among therapists and the situations causing them would probably be experienced as difficult by most, if not all, therapists who are of equivalent theoretical persuasion and at a similar point in their professional development. It follows, that transient difficulties are neither idiosyncratic nor rooted in the lasting personal characteristics of the therapist experiencing them. As their name implies, they are of a transitory, impermanent nature.

The concise definition of transient difficulties, as supplied to raters, was as follows:

These are difficulties in which the situation encountered exposes deficiencies in the therapist's knowledge, technical skills or experience. Though they may be troublesome, vexing, or irritating at the time of their occurrence, they are essentially impermanent in nature and are potentially capable of being remedied through further training and experience. They are likely to be found difficult by any therapist with similar levels of knowledge, technical skills, and experience. They do not reflect the enduring personal characteristics of particular therapists.

4.2.1.2 Paradigmatic Difficulties

In contrast, the central defining feature of paradigmatic difficulties is their link to stable, distinctive personal attributes of the therapists who describe them. Depending on one's frame of reference, such attributes may be understood in terms of intrapersonal qualities, such as therapists' personality structures or significant intrapsychic conflicts, or interpersonal features, such as fixed social response patterns or preferred interactional positions. These characteristics are invoked by the situation described and are crucial to the experience of that situation as difficult. Most other therapists at a similar point in their professional development would not have a difficulty with the same situation. Paradigmatic difficulties are therefore idiosyncratic - typical of individual therapists rather than of situations or of developmental levels. Their enduring nature implies that acquisition of further skills or accumulation of more clinical experience might at best attenuate them, or help the therapist to cope with them, but would not eliminate them. Therapists would need to undergo comprehensive personal change to rid themselves of these difficulties.

The concise definition of paradigmatic difficulties, as supplied to raters, was as follows:

> These are difficulties which arise out of the enduring characteristics of the therapist experiencing them. They may be coped with, accommodated to, or somewhat modified over time, but they are essentially stable in nature. They are idiosyncratic and may be attributed to the therapist's internal conflicts, interpersonal style, or habitual ways of reacting. Their relatively unchanging character makes them typical of (sometimes even prototypical for) a particular therapist, and the situation that evokes them would not be expected to cause similar difficulties for therapists in general. It would require far-reaching personal change for the therapist to become free of such difficulties.

4.2.1.3 A new Category - Situational Difficulties

During the process of refining the above definitions, by piloting them with a number of difficulty accounts from a previous study, it became apparent that there are some difficulties which cannot be considered as paradigmatic, as they do not reflect therapists' idiosyncrasies, but which nevertheless are of a permanent nature. Although one might expect them to be somewhat alleviated, one would not expect them to be erased by further training or by the accumulation of knowledge and experience. Examples of such difficulties might be those which in the original taxonomy had been subsumed under the label of 'painful reality' (situations which are saddening or distressing to the therapist but which cannot be changed and therefore have to be accepted for what they are) or those which involve 'difficult patients' (for instance

those often described as 'borderline') who will cause difficulties for therapists at all levels of professional development. As these difficulties are intrinsic to the situation which is experienced as difficult, the label 'situational difficulties' seemed apposite. In terms of the difficulty factors emerging from the ISPDP (see 2.2.1.2.2 above) these difficulties straddle factors two and three ('Bad Patient' and 'Bad World') whereas transient and paradigmatic difficulties are elaborations of factor one ('Bad Therapist') problems. A consequence of adopting this new category for the rating study was that the original plan of having raters locate difficulty descriptions on a transient - paradigmatic continuum had to be abandoned in favour of independent ratings of all three categories. These were subsequently operationalised as judgements about the 'salience' of each difficulty type for the account provided by therapists.

The concise definition of situational difficulties, as supplied to raters, was as follows:

> These are difficulties which are inherent in the situation encountered by the therapist. They would probably be experienced by most therapists encountering the situation regardless of their levels of knowledge, technical skills or experience. They are not reflective of the therapist's enduring personal characteristics and though they may be attenuated, they cannot be eliminated through further training and experience

4.2.2 The Significance of Contextual Information

Therapists experience difficulties within the context of their own personal and professional development. Some of these contextual features are discernible from the written narratives, others can be gleaned from other sections of TDQ1. Information about length of practical experience and theoretical orientation is likely to help in the

assessment of the generalisibility of a difficulty to other therapists at a similar stage in professional development and so aid in distinguishing between transient and paradigmatic difficulties. The sex of the therapist reporting the difficulty is not always clear from the narrative but might have a bearing on the understanding of the experience. Consequently, information about these three variables was included in the rating materials as far as it was available. As TDQ1 had omitted to ask for length of practical experience at the time when the difficulty occurred, this information was unavailable for past difficulties, whereas for current or very recent difficulties, practice length at the time of completing the questionnaire (which was available) could be taken as a reasonable approximation.

In order to assess the influence of contextual information on the reliability of estimates of the salience of difficulty types, a subset of accounts was initially presented to raters without such information. Having made their judgements, raters were then supplied with the contextual information and invited to revise their ratings if this made a difference to their evaluation. In general, one would expect interrater reliabilities to be higher for the ratings made on the basis of accounts presented with contextual information.

4.2.3 Can Experienced Therapists guess Contextual Information accurately?

If contextual information is helpful in the reliable attribution of difficulty types to narratives, the question arises whether experienced clinical raters might implicitly use such information even when it is not supplied. If so, one might expect raters to be able to infer contextual variables from the narratives alone. While this question is not directly relevant to the questions under study, it held enough interest in itself to prompt a subsidiary investigation. For this reason, a subset of difficulty descriptions

was presented to raters without contextual information and with an invitation to make a guess as to the sex, theoretical orientation(s) and experience level of the therapist who had provided the account. In the event, the data generated by this procedure were not analysed for this thesis and await future exploration.

4.2.4 Ratings of Difficulty Types: Reliability Study

Prior to investigating the correlates and the validity of ratings of type of difficulty, a reliability study of these ratings was carried out. The author served as one rater, along with several external judges. Following Lambert & Hill's (1994) recommendation, it was decided to have all difficulty accounts evaluated by external judges rather than rely on a reliability check of a sample of accounts.

4.2.4.1 The Question of Rater Training

The customary procedure for establishing the reliability of a scale or scoring procedure is to select a number of raters carefully, paying attention to their background and attempting to maximise their suitability and motivation (Moras & Hill, 1991), and then training them until they reach a pre-set criterion. Elliott (1989) suggests aiming for a coefficient alpha of at least .70 for three to four raters before engaging them in the task. Although this kind of procedure was eventually adopted for the evaluation of written accounts in Survey II (see section 5.2.4.1 below), it was decided <u>not</u> to train raters for this earlier stage of the study. The reasons were threefold:

a) using a training procedure might obscure ambiguities in the rating definitions and instructions, because it is not necessarily apparent whether 'trained' raters make their judgements on the basis of the written materials alone or use additional criteria acquired in the training process;

b) the rating procedure was designed to provide information about some supplementary questions (see below), the responses to which might be contaminated by the training process; and

c) the selection of raters was guided by the aim to gather information about the possible influence of professional expertise and amount of clinical experience. Such information might be helpful for the preparation of a "stand-alone" rating manual, as it would give an indication of the importance of having experienced clinicians as raters to other researchers wishing to utilise these measures.

4.2.4.2 Preparation of Rating Materials and Instructions

Instead of employing a training procedure, written instructions were prepared (see appendix 5), designed to provide raters with all the information they needed in order to make reliable judgements. These related to rating booklets (see appendices 6 and 7) which presented one account per page together with a template on which raters could record their judgements (usually on the same page but in some instances - where accounts were exceptionally long - on the following page). This arrangement was chosen in order to reduce raters' opportunities of making recording errors which might have been increased had they, for instance, had to enter their ratings into score sheets. In booklet 2 only, supplementary information and a second template for revised ratings were provided on the back of each page.

Difficulty accounts were arranged into two groups: Batch1 contained an array starting with the 'current difficulty' of respondent no.1, followed by the 'past difficulty' of respondent no.2, the 'current difficulty' of respondent no.3 and so forth, whereas Batch2 had the obverse arrangement, starting with the 'past difficulty' of respondent no.1, followed by the 'current difficulty' of respondent no.2, the 'past difficulty' of

respondent no.3 etc. This produced two arrays in which each account was drawn from a different therapist, thereby avoiding dependencies between cases, which might confound some statistical analyses, and providing a design in which findings could be 'replicated' between the two halves (batches) of the total data set.

Verbal accounts were in general quite concise, with a mean length of about 130 words, but considerable variation, ranging from 7 to 456 words. It seemed therefore important, especially for the shorter accounts, to preserve some contextual information. Consequently, indents and unusual positionings of text, which had been omitted in the text files, were restored for inclusion in the rating booklets as they might provide additional clues to raters about emphases and structures in the written accounts. This was deemed to be the best compromise between providing raters with a facsimile of each account, leaving them either to struggle with handwritings or with information overload if typed transcripts had been provided alongside, and giving them a standardised but sanitised version of the texts which would have left out the writers' idiosyncrasies.

Each difficulty description was supplemented by information about the sex, length of practical experience and theoretical orientation of the therapist rendering the account. In booklet 1 the supplementary information immediately followed the narratives of the difficulties, with length of practical experience suffixed with 'NOW' in the case of past difficulties to remind raters that for these no information was available about the therapists' practice duration at the time when the difficulty had occurred. In booklet 2 the supplementary information was withheld until raters had recorded their initial 'context-free' judgements (for past difficulties) or recorded their guesses as to the sex, practice length and orientation of the therapist providing the account (for current difficulties). It was then supplied for the 'final' ratings.

Following a general introduction to the task and some items asking for general information about the rater (see appendix 6), but preceding the accounts to be rated, was a set of five 'practice accounts' drawn from those questionnaires which were discarded because of missing data or inconsistency of responses to the Familiarity and Impact scales. Raters were directed to consult with the author before proceeding to the main task if they had encountered any ambiguities in the instructions, called to mind any questions not covered by the instructions or experienced any other problems.

The materials supplied with the rating booklets (see appendix 5) consisted of

a) the definitions of the 'transient', 'situational' and 'paradigmatic' difficulty types,

b) the headings and instructions with which accounts had been elicited in TDQ1, and

c) the 'Directions for Raters' in two versions specific to booklets 1 and 2.

The latter explained the distinction between past and current difficulties, introduced the nature and occurrence of supplementary information and acquainted raters with the transcription conventions. Raters were then instructed to use the whole account - including the therapist's coping efforts and the eventual outcome of the situation - for their judgements and were given operational definitions of the anchor points of the four-point salience scale they were to employ for their ratings. These were modifications of the presence/absence rating definitions employed in the original difficulties taxonomy (Binns et al., 1986) which had used a 'definitely present', 'probably present but possibly absent', 'probably absent but possibly present' and 'definitely absent' spectrum to define anchor points.

For the present study, raters were asked to make a judgement about the 'salience' of a category for an account - in effect, to estimate the degree of 'goodness of fit between

narratives and difficulty definitions. The four points of the 0-3 rating scale were labelled in ascending order

0	*not salient*	for this account
1	*possibly salient*	for this account
2	*probably salient*	for this account
3	*almost certainly salient*	for this account

The 'almost' in the label of the last point was meant to acknowledge that, given the level of inference required for the judgements, raters could not be expected to be entirely certain about the salience of a category for an account.

Definitions for the scale points suggest that raters make two different kinds of judgement. The first is concerned with the amount of evidence contained in each account for the salience of each category. The definitions for scale points start accordingly as follows

0 there is *no evidence to suggest* that this category *would apply* to the kind of difficulty encountered by the therapist.

1 there is *little or no direct evidence to suggest* that this category *would apply* to the kind of difficulty encountered by the therapist.

2 there is *clearly some direct evidence to suggest* that this category *would apply* to the kind of difficulty encountered by the therapist...

3 there is *ample direct evidence to demonstrate* that this category *applies* to the kind of difficulty encountered by the therapist.

The focus for this judgement is therefore on the situation as recapitulated by the therapist, with the rater adopting an outsider's perspective and not going beyond what the therapist reports. This is in contrast to the second kind of judgement required,

which is concerned with the plausibility of a category being applicable to a difficulty account, given the situation and the way in which it is described. This latter judgement is only invoked for the first three scale points where salience is lacking or uncertain. The relevant parts of the definition read as follows:

0 *Furthermore*, the situation described and/or other features of the therapist's account make it *somewhat implausible* for the difficulty to have come about in the way the category describes.

1 *However*, the situation described and/or other features of the therapist's account make it *quite plausible* for the difficulty to have come about in the way the category describes.

2 ...*and* the situation described and/or other features of the therapist's account make it *quite plausible* for the difficulty to have come about in the way the category describes.

Unless salience is judged to be 'almost certain', raters are thus invited to adopt a first person perspective, to put themselves in the shoes of the therapist experiencing the difficulty, and to make a plausibility judgement as a result of this identification.

Finally, if there is direct evidence for the salience of a category (scale points 2 and 3), raters are directed to make a judgement about the level of inference required. The last parts of the definitions for these points are therefore as follows:

2 However, the therapist's account *is not* sufficiently explicit or unequivocal to put the relevance of the category beyond doubt; salience *has to be inferred rather than being self-evident*.

3 The account *is* sufficiently explicit or unequivocal to put the relevance of the category beyond doubt; salience *is manifest rather than implied*.

Table 4:1 below summarises the labels and components of salience judgements for each scale point:

No.	Label	Evidence	Plausibility	Inference
0	not salient	no evidence	implausible	
1	possibly salient	little or no direct evidence	quite plausible	
2	probably salient	clearly some direct evidence	quite plausible	required
3	almost certainly salient	ample direct evidence		not required

Table 4:1 - Components of Salience Judgements for Difficulty Types

A decision making sequence for salience ratings might therefore run as follows:

Decide if there is some **clear direct evidence** for the applicability of a category;

if **no**, appraise *plausibility* of salience,

if *implausible*, assign scalepoint '**0**', 'not salient',

if *quite plausible*, assign scalepoint '**1**', 'possibly salient';

if **yes**, appraise necessity for *inference*,

if salience *has to be inferred*, assign scalepoint '**2**', 'probably salient',

if salience *is manifest*, assign scalepoint '**3**', 'almost certainly salient'.

It is evident from this sequence (which is a post-hoc construction and did not form part of the rating directions) that the plausibility part of the definition for scalepoint 2 is redundant. It had originally been included in order to maintain continuity between definitions, with each formulation building on the composition of the previous one.

Raters were, however, furnished with some additional suggestions to help them in forming judgements. The first three of these - advising them to take into account therapists' emotional reactions and coping efforts and whether the situation was resolved or not - are mainly amplifications on the instruction to read the whole account, making explicit that all information can potentially contribute to judgements,

although focusing on resolution goes somewhat beyond simple situation outcome and suggests an additional criterion for transient difficulties. The remaining four suggestions are in the form of questions which raters could ask themselves. Three of these,

- Is this situation or the therapist's reaction to it within the range I would find reasonable or would have predicted?
- Would I expect the therapist to react similarly to a situation outside therapy?
- How would another therapist react in this situation?

amplify therapists' idiosyncratic reactions to difficulties and were thought to be of help in distinguishing paradigmatic difficulties, whereas the fourth,

- If the therapist possessed additional knowledge, technical skills, or experience, would he/she still find the situation difficult?

was intended to provide an additional pointer to transient difficulties. However, at this point not enough conceptual clarity had been achieved to formulate explicit rating instructions from these markers and they were left as points for general consideration.

Apart from the salience ratings, raters were also instructed to make a global judgement about which single category would best fit an account. A space for recording this 'single choice' rating was provided as part of the template on each page.

Finally, in the directions supplied with booklet 2 only, raters were instructed to guess the a) sex, b) level of experience and c) theoretical orientation of the therapist who had supplied the difficulty account and to record their guesses in a template before rating salience and single choice. Options on this template consisted of

a) Male / Female

b) less than 6 years / 6 - 12 years / more than 12 years and

c) Psychodynamic / Behavioural / Cognitive / Systemic / Humanistic.

The options in b) were derived from the distribution in the sample collected with TDQ1 and represent approximately one third of respondents each, the options in c) represent the categories in TDQ1 which had formed the basis for therapists' self-descriptions. Having recorded their initial salience and single choice judgements, raters were then directed to take note of the actual supplementary information shown on the reverse of each page and to enter their modified ratings on a second template or to re-enter their original ratings if the contextual information had not caused them to change their minds.

4.2.4.3 Selection of Raters

The selection of a pool of raters followed the strategy of employing a spectrum of judges, ranging from those most familiar with the topic to those with a basic understanding of psychological therapies but without practical experience as therapists or supervisors. This, it was hoped, would provide evidence both for the clarity and comprehensiveness of the rating instructions and for the hypothesis that the reliability of ratings would improve with increasing clinical experience. Located at one end of the spectrum were the author (R1) and the thesis supervisor (R2), who one would expect to be most knowledgeable about the concepts employed, but also most likely to arrive at ratings on the basis of understandings and information assimilated without conscious awareness and therefore not contained in the rating instructions. At the opposite end of the spectrum was a research assistant (R3), who approached the rating task relatively naively, having only just begun a training in

counselling skills and having no experience as a psychological therapist or as a supervisor. Located at intermediate points of the spectrum were

a) two members of the original collaborative research group involved in developing the difficulties taxonomy; both were experienced supervisors, one (R4) now mainly working in the areas of Health Psychology and Psychiatric Rehabilitation, the other (R5) now retired from the Health Service but retaining an interest in the field of psychological therapies;

b) a clinical psychologist trained in dynamic psychotherapy (R6), an experienced supervisor, now mainly working in the area of Psychiatric Rehabilitation; and

c) a specialist psychodynamic psychotherapist (R7) from a social work background, also with extensive experience of supervising psychotherapists.

Six of the seven raters were male, one was female.

4.2.4.4 Interrater Reliabilities

Comparisons between the two parallel batches of data (see 4.2.4.2 above) are not straightforward as only four out of seven raters completed both booklets. Some raters may have been deterred by the additional complexity of the second booklet. With hindsight, it might have been better not to have introduced these complications and to have restricted the scope of the investigation.

Because of the diversity of the group of raters, attention has been paid to pairwise as well as to overall reliabilities. Both forms of analysis are reported below for each batch of data in turn. Only salience ratings have been used, single choice judgements were not analysed in this study.

4.2.4.4.1 Batch 1

4.2.4.4.1.1 Pairwise correlations between raters and individual salience means

All seven raters completed the first batch of ratings contained in booklet 1. The pairwise correlations between raters for each difficulty type are displayed in the tables below; salience means and standard deviations for each rater are shown in the diagonal in the format **[MEAN]** (means displayed in square brackets and **bold** SD typeface in the upper line).

TRANSIENT DIFFICULTY SALIENCE:

	R1	R2	R3	R4	R5	R6	R7
R1	**[1.19]** .94						
R2	**.41**	**[.98]** 1.04					
R3	**.24**	**.20**	**[1.67]** .96				
R4	**.15**	**.28**	**.01**	**[1.24]** .92			
R5	**.47**	**.56**	**.34**	**.34**	**[.93]** .89		
R6	**.35**	**.47**	**.34**	**.27**	**.48**	**[1.61]** .59	
R7	**.16**	**.18**	**.31**	**.17**	**.27**	**.42**	**[1.66]** .99

Table 4:2 - Pairwise Correlations between Raters: Transient Difficulty Salience

SITUATIONAL DIFFICULTY SALIENCE:

	R1	R2	R3	R4	R5	R6	R7
R1	**[1.39]** .91						
R2	**.59**	**[1.85]** 1.20					
R3	**.33**	**.26**	**[2.73]** .58				
R4	**.47**	**.29**	**.22**	**[1.46]** .99			
R5	**.38**	**.29**	**.34**	**.25**	**[2.58]** .79		
R6	**.37**	**.35**	**.46**	**.27**	**.20**	**[2.44]** .65	
R7	**.53**	**.50**	**.39**	**.39**	**.35**	**.64**	**[2.27]** .78

Table 4:3 - Pairwise Correlations between Raters: Situational Difficulty Salience

PARADIGMATIC DIFFICULTY SALIENCE:

	R1	R2	R3	R4	R5	R6	R7
R1	**[1.86]** .84						
R2	**.72**	**[1.49]** 1.17					
R3	**.53**	**.44**	**[.76]** 1.01				
R4	**.38**	**.39**	**.22**	**[.90]** .82			
R5	**.66**	**.65**	**.51**	**.46**	**[1.64]** .94		
R6	**.35**	**.44**	**.39**	**.52**	**.42**	**[1.85]** .69	
R7	**-.09**	**.14**	**.24**	**.08**	**.07**	**.09**	**[.34]** .55

Table 4:4 - Pairwise Correlations between Raters: Paradigmatic Difficulty Salience

For the transient difficulty type, R3 shows the highest mean salience rating. This is consistent with the expectation that someone at the beginning of counselling training would perceive most difficulties in terms of skills deficits or situational factors. Somewhat more surprising are the nearly equally high means for raters R6 and R7. For R7 a possible explanation may derive from his particular perception of the concepts underlying the difficulty types (which is reported in detail below), but for R6 no such explanation suggests itself. Average pairwise correlations per rater range from .20 to .41, the lowest are achieved by R4 (.20), R3 (.24) and R7 (.25). It is also noteworthy that in a descending rank order of pairwise correlations, the agreement between R1 and R2 only ranks as number six which provides some reassurance that,

at least for this difficulty type, agreement between raters is more determined by the rating instructions than by prior knowledge. It might be argued, however, that the concept of transient difficulties is not unidimensional in this set of ratings; in other words, that discrepancies in the ratings might be attributable to divergent understandings of the concept, each of which is shared by a subgroup of raters, rather than to rater error. If this was the case, one would expect a factor analysis of the whole set of ratings to yield factors which are clearly interpretable as different but consistent understandings of the concept. A principal component analysis with subsequent varimax rotation of the whole set of 413 (7 raters x 59 accounts) salience ratings for transient difficulties did indeed extract two factors with eigenvalues > 1. These differentiate between the subgroups of (R1, R2, R4, R5) and (R3, R7), while the loadings for R6 are almost equally distributed between the two factors. While there is therefore evidence that R3 and R7 share an understanding of the transience concept which is divergent from most of the other raters, a closer examination of R7's decision-making (reported below) suggests that the discrepancy should be interpreted as rater error rather than as an alternative construction of transient difficulties.

For the situational difficulty type, R3 again shows the highest mean salience rating, in line with expectations. Average pairwise correlations per rater range from .30 to .45, with the lowest achieved by R5 (.30) and R4 (.32). The agreement between R1 and R2 is the second highest of all. Principal component analysis of all situational salience ratings show the concept to be consensually understood in this set of ratings as only one factor with an eigenvalue > 1 could be extracted.

For the paradigmatic difficulty type, R7 shows by far the lowest mean salience rating, which is surprising given his level of experience. The next lowest mean shown by R3, however, is again in line with expectations. Average pairwise correlations per rater

range from .34 to .46, with the spectacular exception of R7 who only achieves .09. The agreement between R1 and R2 is the highest of all, raising a concern that agreement for this difficulty type may be mainly attributed to prior familiarity with the concepts. There are, however, a number of high pairwise correlations between other raters and the question is further addressed empirically below in the section where agreement is calculated in terms of Cronbach's alphas. Principal component analysis of all paradigmatic salience ratings shows that two factors with an eigenvalue > 1 can be extracted. The second factor, however, is entirely accounted for by rater 7 and the concept of paradigmatic difficulties can therefore be regarded as consensually understood in this set of ratings.

In order to understand better why rater 7 should diverge so unexpectedly from the other raters, especially in respect of paradigmatic difficulties, his perceptions were explored in an extended debriefing session. The following issues emerged in the interview (which lasted approximately 45 minutes):

Rater 7 had an understanding of paradigmatic difficulties as being somewhat disreputable for trained psychotherapists, who he thought could be expected to have addressed such enduring personal issues in their training therapies. He had therefore experienced a tension about the definition of paradigmatic difficulties as given in the rating instructions, but had not sought to resolve this - as requested - after rating the five 'practice' accounts because he had undertaken the rating task at a time when the author was on leave and not easily approachable. R7 spoke of his being impressed by the extent and severity of the struggles which the contributors of the accounts reported and of consequently feeling that giving high paradigmatic salience ratings to their accounts would be 'disloyal' and tantamount to trivialising their efforts. He had also been a contributor to the study and, having rated his own accounts as 'non-

paradigmatic' felt that others deserved the same consideration.[11] Having his attention drawn again to the definition of paradigmatic difficulties, he thought that he might have misinterpreted this category. He felt clear about the attribution of difficulties to situational factors (locating them in the patient or in external circumstances) which made it plausible that difficulties correctly located in therapists, but erroneously not attributed to their enduring characteristics ('paradigmatic'), would of necessity be attributed to skills deficits ('transient') by him, accounting for his high mean salience ratings for this difficulty type.

Debriefing of other raters revealed that R4, having undertaken her ratings on a train journey, had quickly tired of the task, feeling that she had underestimated (or been misled about) the effort required in completing it, and ending up rather resentful about having agreed to it. This may go some way towards accounting for her low level of agreement with other raters which is otherwise surprising given her training and experience. Raters 2, 5 and 6 reported experiencing intermittent doubts about the accuracy and reliability of their ratings but encountered no other problems.[12]

4.2.4.4.1.2 Cronbach's alphas for the whole group of raters

In order to a) provide an estimate of reliability for the whole group of raters and b) identify subsets of raters which achieved the highest reliabilities among themselves for each of the difficulty types (so that subsequently 'transient', 'situational' and 'paradigmatic' scores, derived from the mean of the ratings of the most reliable subset, could be assigned to each account), coefficient alphas were calculated, thereby treating each rater as if s/he was one item contributing to (or subtracting from) the

---

[11] In terms of established sources of unreliability in human judgements, R7 exhibited a particular form of 'leniency bias' (Saal, Downey, & Lahey, 1980).

[12] Rater1 felt enthusiastic about the task and confident about his ratings - but then he would, wouldn't he.

reliability of scales measuring the three difficulty types. Results are shown in Table 4:5 below. However, as Shrout and Fleiss (1979) observe, the use of coefficient alpha is only indicated where judges can be regarded as a a fixed factor - the only raters which are of interest. It may therefore be appropriate for determining agreement between raters within this particular reliability study. The argument that difficulty types can generally be rated reliably (on the basis of their definitions and the rating instructions), however, requires judges to be treated as a random factor - one sample out of the population of possible raters. With hindsight, it would therefore have been preferable to use the intraclass correlation coefficient referred to as ICC(2,k) by Shrout and Fleiss (1979), which takes account of variability between judges.

DIFFICULTY TYPE :	α All Raters (R1 - R7)	α (R1, R2, R5, R6)	α Best Subset	α Best Subset excl. R1	α Best Subset excl. R1, R2
TRANSIENT	**.74**	**.76**	**.76**	**.71**	**.66**
SITUATIONAL	**.80**	**.69**	**.80**	**.74**	**.72**
PARADIGMATIC	**.81**	**.82**	**.85**	**.79**	**.73**

Table 4:5 - Interrater Reliabilities for Batch1

Column 1 shows Cronbach's alphas calculated from the whole group of seven raters for each of the three difficulty types. In column 2 reliabilities are shown for a group of raters (R1, R2, R5, and R6) which excludes those which consistently had the lowest pairwise correlations with other raters in the set. For transient and paradigmatic difficulty types this results in a slight improvement, whereas for situational difficulty saliences all seven raters contributed positively to the overall reliability and excluding some of them depresses the alpha coefficient. Column 3

shows the reliabilities for a subgroup of raters selected to maximise alphas. For transient difficulties this is the same as in column 2, for situational difficulties it is coterminous with the whole group of seven raters and for paradigmatic difficulties excluding another rater (R6) from the set in column 2 results in a further improvement of coefficient alpha. The mean rating for each difficulty type of the subgroups shown in column three is used for further analyses. The magnitudes of alphas, ranging from .76 to .85, is encouraging, especially when considering that this is derived from an untrained group of heterogeneous raters with widely differing levels of expertise and motivation.

It might be argued, however, that these reliabilities are inflated because of the close familiarity of R1 and R2 with the concepts rated in this study, a concern already raised above in connection with pairwise correlation between these two raters. In order to test out empirically whether this was the case, reliabilities were calculated for subsets selected to maximise alphas after excluding R1 (in column 4) or after exclusion of R1 and R2 (see column 5). The results show alphas on the latter, more stringent, criterion ranging from .66 to .73, providing reassurance that reliable ratings were possible on the basis of the written instructions alone, without the necessity of recourse to additional criteria derived from familiarity with the topic under investigation.

4.2.4.4.2 Batch 2

Only four raters (R1, R2, R3 and R6) completed the second batch of ratings contained in booklet 2. R4 had selected herself out of further participation on the grounds that she could not spare the time and did not feel particularly motivated to continue with the ratings. R5 had taken a number of months to complete booklet 1, having been

absorbed by the arrival of a new-born in his family and could not guarantee to complete booklet 2 within the time frame required by this study and R7 was excused from taking part in further ratings because of his idiosyncratic interpretation of the categories. It is possible that the more complex requirements of half of the ratings in booklet 2 contributed to the reluctance of two of the raters to carry on. The general experience of the difficulties of obtaining ratings within a reasonable timescale and the sense that some of the raters had undertaken part of the task 'on sufferance' led to the decision to pay raters for the subsequent study as described in 5.2.4.1 below.

4.2.4.4.2.1 Pairwise correlations between raters and individual salience means

The pairwise correlations between raters for each difficulty type are shown in the following three tables. As before, salience means and standard deviations for each rater are shown in the diagonal in the format **[MEAN]** (means displayed in square SD brackets and **bold** type in the upper line).

TRANSIENT DIFFICULTY SALIENCE:

	R1	**R2**	**R3**	**R6**
R1	**[1.71]** .68			
R2	**.45**	**[.88]** 1.01		
R3	**.28**	**.29**	**[1.11]** .73	
R6	**-.02**	**.12**	**.38**	**[1.86]** .44

Table 4:6- Pairwise Correlations between Raters: Transient Difficulty Salience

SITUATIONAL DIFFICULTY SALIENCE:

	R1	R2	R3	R6
R1	**[1.38]** .96			
R2	**.69**	**[1.66]** 1.07		
R3	**.42**	**.37**	**[1.82]** .58	
R6	**.44**	**.43**	**.38**	**[2.45]** .54

Table 4:7 - Pairwise Correlations between Raters: Situational Difficulty Salience

PARADIGMATIC DIFFICULTY SALIENCE:

	R1	R2	R3	R6
R1	**[1.57]** .78			
R2	**.53**	**[1.80]** 1.02		
R3	**.00**	**.05**	**[.46]** .76	
R6	**.32**	**.34**	**-.02**	**[1.79]** .56

Table 4:8 - Pairwise Correlations between Raters: Paradigmatic Difficulty Salience

For the transient difficulty type, R6 shows the highest mean salience rating, closely followed by R1. For R6 this is consistent with his ratings of batch1 and is probably best understood as an idiosyncratic baseline. R3 does not show the same elevation of

means as in batch1 which weakens the argument that readiness to perceive difficulties as transient may be inversely related to experience. Average pairwise correlations per rater range from .16 (R6) to .32 (R3), somewhat lower than in batch1. Principal components analysis shows that two factors with an eigenvalue > 1 can be extracted, however, the second factor is entirely accounted for by R6 whose low agreement with the group mean eventually led to his exclusion from the subset of judges used to calculate mean ratings and reliabilities for this difficulty type.

For the situational difficulty type, R6 shows a mean salience rating practically identical to that for batch1; the most notable difference from the previous ratings is shown in the mean for R3, which is substantially lower. Average pairwise correlations per rater range from .39 (R3) to .52 (R1), somewhat higher than those in batch1. Principal component analysis of all situational salience ratings shows the concept to be consensually understood in this set of ratings as only one factor with an eigenvalue > 1 could be extracted.

For the paradigmatic difficulty type, R3 (who does not have experience as a clinician) shows again by far the lowest mean salience ratings and his pairwise correlations, averaging .02, reveal his judgements to be entirely unrelated to that of other raters, whose salience means are broadly in line with each other and with those recorded for batch1. The other three average pairwise correlations per rater range from .21 to .31, clearly lower than those in the first batch. Not surprisingly, principal component analysis of all paradigmatic salience ratings shows that R3's ratings form a factor of their own (with an eigenvalue of 1.001), whereas the other factor with an eigenvalue > 1 accounts for the other three raters. This suggests that R3 has an idiosyncratic understanding of the paradigmatic difficulty concept, whereas the understanding of the other three raters appears to be consensual.

4.2.4.4.2.2 Cronbach's alphas for the whole group of raters

As for batch1, coefficient alphas (Cronbach, 1951) were calculated in order to estimate reliabilities for the whole group of raters and to identify subsets of raters which maximised reliabilities for each of the difficulty types as a basis for the computation of mean 'transient', 'situational' and 'paradigmatic' salience scores. Results for batch2 and comparisons with batch1 are shown in Table 4:9 below:

INTERRATER RELIABILITIES (BATCH2):

	α All Raters	α Best Subset	α corrected	Best α Batch1
TRANSIENT	**.57**	**.60**	**.67**	**.76**
SITUATIONAL	**.76**	**.76**	**.85**	**.80**
PARADIGMATIC	**.51**	**.68**	**.76**	**.85**

Table 4:9 - Interrater Reliabilities for Batch2

Column 1 shows alphas calculated from salience ratings of the whole group of four raters for each of the three difficulty types. In column 2 reliabilities are shown for a subgroup of raters selected to maximise alphas. For situational difficulty saliences this is the same as in column 1, as all four raters contributed to the overall reliability. Reliability was improved for transient difficulty saliences by excluding the ratings of R6, and for paradigmatic difficulty saliences by excluding those of R3 and R6. The mean rating for each difficulty type of the subgroups used for column 2 was subsequently employed in further analyses. The magnitudes of alphas, ranging from .60 to .76, are substantially smaller than those in the first batch which are shown in column 4. As this difference may be partly accounted for by the smaller groups of raters on which the results are based, column 3 shows alphas corrected for group size

by using the Spearman-Brown formula (Nunnally, 1978), estimating reliabilities for groups of four ('transient'), seven ('situational') and three ('paradigmatic') respectively. While this correction raises the reliability estimate for situational difficulty saliences to a level above that of the first batch, the corrected alphas for the other two difficulty types are still some way short of those obtained for batch1. Apart from random fluctuations it seems likely that the format of rating booklet 2 also introduced an error by complicating the rating. For R6, who, having been recruited late to the task, completed all ratings in a single day (after first filling in the subsequent questionnaire - TDQ2), a significant fatigue effect also seems likely. Even so, the lower reliability estimate for 'paradigmatic' salience ratings remains disappointing and to some extent unexplained.

4.3 Results

Having estimated the reliabilities for each of the difficulty types in both batches, mean salience ratings for each difficulty account are now available for use in further analyses. Two investigations are of interest here, a) the relationships of difficulty types to each other and b) their associations with the FID scales.

4.3.1 Difficulty Types - Intercorrelations

Intercorrelations between difficulty types were calculated and subsequently corrected for attenuation (by dividing each correlation by the product of the square roots of the associated reliabilities (Lord & Novick, 1968)) to give a better indication of discriminant validity (Schmitt, 1996). The two tables below display intercorrelations for each batch of data. Reliabilities for each difficulty type (expressed as Cronbach's alphas) are shown in square brackets in the diagonal, uncorrected correlations in

italics below the diagonal and corrected correlations in normal script above the diagonal.

BATCH 1 :

N = 59	TRANSIENT	SITUATIONAL	PARADIGMATIC
TRANSIENT	**[.76]**	**-.50**	**.05**
SITUATIONAL	***-.39***	**[.80]**	**-.61**
PARADIGMATIC	***.04***	***-.50***	**[.85]**

Table 4:10 - Intercorrelations between Difficulty Types for Batch1

BATCH 2 :

N = 56	TRANSIENT	SITUATIONAL	PARADIGMATIC
TRANSIENT	**[.60]**	**-.48**	**.08**
SITUATIONAL	***-.32***	**[.75]**	**-.84**
PARADIGMATIC	***.05***	***-.60***	**[.68]**

Table 4:11 - **Intercorrelations between Difficulty Types for Batch2**

There is a consistent picture emerging from both data sets, indicating that transient and paradigmatic salience ratings are independent of each other, but are both substantially negatively associated with situational salience. It appears that raters tended to locate difficulties either inside the therapist (transient and paradigmatic salience) or outside the therapist (situational salience). This result contributes to the construct validity of the 'situational difficulty' concept which had been devised to accommodate difficulty accounts that did not easily fit into the transient and paradigmatic categories (see 4.2.1.3 above). Once the decision had been made to locate the difficulty mainly inside the therapist, raters' judgements about the source of the difficulty - whether it was based on skills deficits (transient salience) or

therapists' enduring characteristics (paradigmatic salience) - were made independently of each other. This result supports the decision to treat transient and paradigmatic difficulties as independent dimensions, rather than opposite ends of the same construct.

4.3.2 Associations between Difficulty Types and FID Scales

The analysis of the correlations between difficulty type mean salience scores and the indices for Familiarity, depth of emotional Impact and Duration offers an opportunity of investigating their hypothesised relationships (see 1.3, research question 2 above) empirically. Because of the way in which data batches were constructed (see 4.2.4.2 above), each array contains equal numbers of 'past' and 'current' difficulty accounts and of the associated versions of the FID scales. Table 4:12 below shows the associations of Familiarity, Impact and Duration scale scores and the mean salience scores for the three difficulty types across both data batches.

	MEAN SALIENCE SCORES					
	TRANSIENT		**SITUATIONAL**		**PARADIGMATIC**	
	BATCH1	BATCH2	BATCH1	BATCH2	BATCH1	BATCH2
Familiarity	-.35	.02	-.04	-.18	-.02	.18
Impact	.08	-.09	.09	.09	.13	.25
Duration	-.03	-.15	.16	.01	.23	.25

Table 4:12 - TDQ1: Correlations between Difficulty Types and FID Scales

The only consistent associations occur between paradigmatic salience and Impact and Duration. Both are in the expected direction and appear to provide some support for the postulated relationship between paradigmatic difficulties and high FID scores. The results need to be treated with some caution, however, because of the

combination of past and current difficulties in each batch mentioned above. Table 3:10, in the previous chapter, indicated that the past and current versions of the Familiarity scale may behave quite differently, which may account for the absence of consistent relationships between them and the difficulty types. Clarification of the issue may be found in data from TDQ2 (see 5.3.3.4 below) which are not subject to the same caveat.

4.4 Difficulty Types and Taxonomy Ratings: Validity Study

Some of the raters involved in the reliability study reported in 4.2.4 above had also been asked to rate the accounts collected with TDQ1 using the taxonomy of therapist difficulties (see 1.2 above). The relationships between taxonomy categories and difficulty types are of interest because they potentially provide evidence for the construct validity of the latter, and for the conceptual coherence of both research strategies. Four judges supplied ratings for the first batch of data and three judges for the second batch. After excluding the least reliable rater for each batch, satisfactory reliabilities were obtained, with three-rater alphas for batch1 ranging from .65 to .89 and two-rater alphas for batch2 ranging from .65 to .86.

Table 4:13 below shows product-moment correlation coefficients between mean presence/absence ratings for taxonomy categories (Binns et al., 1986) (see appendix 1) and mean salience scores for difficulty types. Figures in square brackets, below the taxonomy category and difficulty type labels, represent reliabilities for batch1 and batch2 respectively. Figures in the cells represent correlation coefficients; those in the top left corners relate to batch1, those in the bottom right corners relate to batch2. Figures displayed in *italics* represent uncorrected correlations, those displayed in **bold** script represent correlation coefficients corrected for attenuation.

N = 59 N=56	**TRANSIENT** [.76 / .60]	**SITUATIONAL** [.80 / .75]	**PARADIGMATIC** [.85 / .68]
Therapist **Incompetent** [.89 / .84]	*.39* **.47** *.14* **.20**	*-.40* **-.47** *.06* **.08**	*.28* **.32** *.08* **.11**
Therapist **Damaging** [.87 / .81]	*.12* **.15** *-.07* **-.10**	*-.21* **-.25** *.07* **.09**	*.28* **.33** *.04* **.05**
Therapist **Puzzled** [.85 / .70]	*.36* **.45** *.11* **.17**	*-.16* **-.19** *.13* **.18**	*.18* **.21** *-.32* **-.46**
Therapist **Threatened** [.85 / .81]	*.04* **.08** *.00* **.00**	*.23* **.28** *.23* **.29**	*.01* **.01** *.02* **.03**
Therapist **Out-of-Rapport** [.75 / .77]	*.22* **.29** *-.02* **-.03**	*.07* **.09** *-.20* **-.26**	*.05* **.06** *.42* **.60**
Therapist's **Personal Issues** [.85 / .57]	*-.15* **-.19** *.03* **.05**	*-.39* **-.47** *-.12* **-.18**	*.70* **.82** *.12* **.19**
Ethical Dilemma [.79 / .86]	*.03* **.04** *-.30* **-.42**	*.12* **.15** *.34* **.42**	*-.06* **-.07** *-.45* **-.59**
Painful Reality [.84 / .66]	*-.43* **-.54** *-.28* **-.44**	*.44* **.54** *.27* **.38**	*-.24* **-.28** *-.08* **-.12**
Therapist **Stuck** [.65 / .65]	*.15* **.21** *-.12* **-.19**	*-.10* **-.14** *.00* **.00**	*.05* **.07** *.20* **.30**
Therapist **Thwarted** [.86 / .76]	*.14* **.17** *-.06* **-.09**	*-.10* **-.12** *-.08* **-.11**	*.17* **.20** *.19* **.26**

Table 4:13 - Correlations between Difficulty Types and Taxonomy Categories

It is apparent that a number of corrected correlations are non-trivial and consistent across batches. If one defines an association as 'noteworthy' if it is above .15 (in absolute terms) and in the same direction across both batches, about one in every three corrected correlations meet that criterion.

Looking along the rows of the table, the clearest picture emerges for the 'Painful Reality' category. As one would expect - given that it should denote difficulties located outside the therapists experiencing them - this category is substantially

positively associated with situational difficulty salience, but negatively related to transient and paradigmatic salience. Therapists' Personal Issues - which one would expect to be located inside therapists and related to personal characteristics - show indeed positive associations with paradigmatic difficulty salience but negative associations with situational difficulty salience.

Looking down columns, it becomes apparent which taxonomy categories might serve as markers for difficulty types. In order to test for statistical significance, combined probabilities for both samples can be estimated with the formula given by Winer (1971), where z is computed by dividing the sum of the t-values from the tests for both samples by the square root of the sum of the degrees of freedom for each test, each divided by the result of subtracting 2 from its value.[13] Given that each difficulty type has been correlated with ten variables (taxonomy categories), it seems reasonable to adopt a criterion of p = .005 (.05/10) as an indicator of statistical significance. As all 'noteworthy' associations were in the expected directions, one-tailed significance tests have been used throughout.

Transient difficulty type is reliably associated with three taxonomy categories[14]:

T- Incompetent (r = .39; df = 57; t =3.14; p = .001 / r = .14; df = 54; t = 1.03; p = .15 / z = 2.90; p(comb) = .001)

T-Puzzled (r = .36; df = 57; t =2.91; p = .003 / r = .11; df = 54; t = .85; p = .20 / z = 2.61; p(comb) = .004)

---

[13] $z = \frac{t_1 + t_2}{\sqrt{\frac{df_1}{df_1 - 2} + \frac{df_2}{df_2 - 2}}}$

[14] In the expressions in brackets, the first set of values (r; df; t; p) relates to batch1, the second set to batch 2, and the third set (z; p(comb)) to the combined probability calculated from both samples.

Painful Reality (r = -.43; df=57; t = 3.54; p < .001 / r = -.28; df=54; t = 2.12; p = .02 / z = 3.93; p(comb) < .0005)

Transient difficulties are therefore positively associated with therapists' negative evaluations of their own performance and with their uncertainties about how to understand a situation (deficits in knowledge and experience) or how best to proceed (skills deficits), but negatively associated with painful circumstances beyond their control (external location).

Situational difficulty type is reliably associated with two taxonomy categories,

Personal Issues (r = -.39; df=57; t = 3.19; p = .001 / r = -.12; df=54; t = .86; p = .20 / z = 2.81; p(comb) = .003)

Painful Reality (r = .44; df=57; t = 3.68; p = .001 / r = .27; df=54; t = 2.02; p = .024 / z = 3.96; p(comb) < .0005)

while the associations with two other categories approach statistical significance:

T-Threatened (r = .23; df=57; t = 1.77; p = .041 / r = .23; df=54; t = 1.72; p = .046 / z = 2.42; p(comb) = .008)

Ethical Dilemma (r = .12; df=57; t = .91; p = .18 / r = .34; df=54; t = 2.68; p = .005 / z = 2.49; p(comb) = .006)

Situational difficulties are evidently positively associated with the "external" categories of painful circumstances beyond the therapist's control and divided loyalties between conflicting demands, while being inversely related to the "internal" category of personal issues intruding into the therapy. There is also an indication that they are positively associated with therapists' feelings that they need to defend themselves against threats, which is likely to be related to 'difficult' patients.

Paradigmatic difficulty type is reliably associated with one taxonomy category,

Personal Issues (r = .70; df = 57; t = 7.35; p < .001 / r = .12; df = 54; t = .86; p = .20 / z = 5.7; p(comb) < .0005)

while the association with another taxonomy category is noteworthy but does not reach statistical significance:

T-Thwarted (r = .17; df = 57; t = 1.31; p = .10 / r = .19; df = 54; t = 1.46; p = .075 / z = 1.92 ; p(comb) = .027)

Paradigmatic difficulties are thus positively associated with factors originating inside the therapist intruding into the therapy. There is also weak evidence that they might relate to therapists' feelings of being impeded or frustrated by their patients which is likely to be related to therapist hostility.

For situational salience a regression model can be fitted reliably across both batches (B1 And B2) as follows[15]:

Situational Salience = Constant + Painful Reality - Personal Issues + T-Threatened + Ethical Dilemma

(B1: Multiple r = .61; F = 8.01; p < .0005 / B2: Multiple r = .55; F = 5.46; p = .001)

The multiple correlations indicate that approximately one third of the variation in situational salience scores can be explained by a linear combination of the four predictor variables.

---

[15] Taxonomy variables had been transformed to reduce skewness of distributions. Scatterplots and influence plots were generated for all variables and were examined for non-linearity of data, possible outliers, or values exerting undue leverage (Lovie, 1991). None of these gave cause for concern. The ratio of cases to independent variables is better than 13:1, lower than some would recommend (Tabachnik & Fidell, 1989), but the replication design is a safeguard against overfitting. High values for 'tolerance' (the inverse of the multiple r of an independent variable with all other predictors) for each predictor variable indicated that multi-collinearity is unlikely to be a problem (Wilkinson, 1996). Normal probability plots of residuals showed these to be approximately normally distributed.

In summary, there is evidence for the conceptual coherence between taxonomy categories and difficulty types, with seven out of ten categories being associated with one or more of the difficulty types. Only the categories of T-Damaging, T-Out-of-Rapport, and T-Stuck lack such associations. Especially the situational difficulty type shows good integration with the previous theoretical framework. In contrast, the paradigmatic difficulty type - being of particular interest in the present study - appears to tap a dimension which, apart from the category of 'therapist's personal issues', has not been covered well by the taxonomy.

4.5 Summary of Results and Discussion

The second study based on the data collected with TDQ1 ensued from the inconclusive results obtained in the first study and led to an extensive reformulation and a theoretical clarification of the concepts concerning the pervasiveness of therapist difficulties. The notions of transient and paradigmatic difficulties endured. However, whereas previously they had been regarded as endpoints of a postulated continuum (see 1.3 above), they are now understood as relatively independent constructs. This revised view has also found empirical support (see 4.3.1 above). Furthermore, the two types of difficulty experiences have now been joined by a third - situational difficulties - which arose out of the practical experience of relating difficulty types to written accounts of therapists' difficulty experiences. Situational difficulties appear to be substantially inversely related to the other two types. This alternative relationship can also be construed in terms of the 'location' of a difficulty (it being perceived as originating either inside or outside the therapist experiencing it) which links it to the results of two previous empirical investigations - the survey of British psychotherapists (Schröder et al., 1987a) and the ISPDP (Schröder, 1997).

Detailed and precise definitions have been elaborated for all three difficulty types and their 'goodness-of-fit' in relation to verbal reports of difficulty experiences has been operationalised in a four-point 'salience' rating scale which is anchored by exact definitions and now also supported by decision-making rules. The difficulty type and scale-point definitions formed the basis for a set of rating instructions which have been used in the reliability study of the ratings of TDQ1 accounts. It has been demonstrated that accounts can be reliably rated on the basis of these instructions by external judges, who have not been specifically trained for this task, provided they comprehend the directions and adhere to them. Experience as a clinician and/or clinical supervisor appears to increase the likelihood of being a reliable rater.

Mean salience scores for the three difficulty types, derived from averaging the scores of a reliable set of raters, provided serviceable dependent variables for further analysis. Their relationships to the Familiarity and Impact scales and to the Duration index are somewhat inconclusive. It appears that Impact and Duration show a positive association with paradigmatic difficulty type as expected, however, Familiarity seems unrelated to any difficulty type. This may be partly due to the format in which familiarity data were collected - two versions of the same scale, one relating to a 'past difficulty', the other to a 'current or very recent' difficulty. This may have made it difficult for respondents to recall accurately how familiar a 'past difficulty' was when it first occurred.

Finally, the relationships of difficulty type salience ratings and taxonomy ratings based on the same accounts support the construct validity of the reformulated and expanded concepts. Particularly situational difficulties show evidence for firm links with taxonomy categories, whereas the paradigmatic type appears to expand the

theoretical understanding of therapists' difficulty experiences into an area not mapped in any great detail by previous theoretical formulations.

In summary, the study reported in this chapter accomplished a refinement of the theoretical understanding of the pervasiveness of therapists' experiences of difficulties in their practice. The revised concepts can be reliably rated and are validated against previous work. They can now be seen as forming part of a 'nomological net' (Cronbach & Meehl, 1955) of interrelated constructs and provide a basis for the empirical study reported in the next chapter.

5. Chapter Five: Survey 2 - The Third Study

Survey 1 and its two related studies had resulted in a method of reliably converting therapists' narratives into salience ratings for each of the three 'difficulty types' (transient, situational, and paradigmatic) which could then serve as dependent variables. It had also succeeded in constructing reliable measures of three experiential concomitants of difficulties (their familiarity, the depth of their emotional impact, and their duration) which could serve as independent variables. Other sets of independent variables measuring 'competency deficits' (for transient difficulties) and 'enduring personal characteristics' (for paradigmatic difficulties) had yet to be devised, and their construction or selection became one major focus for the next study. This third study was based on a fresh survey which made use of a revised and expanded questionnaire and a much broader sampling strategy, but followed the previous method of eliciting written accounts from respondents together with a number of related measures. The new questionnaire was comprehensive and generated more data than could be analysed in the current investigation. Some material therefore awaits future exploration. The rationale for this study, which guided the construction of the new questionnaire, is detailed below.

5.1 Rationale

The main points considered in planning the third study - its links to the preceding studies, the general hypotheses under investigation, and the reasons for a bi-lingual survey - are outlined below. There was, however, one feature which had not been explicitly planned for: The strategy of eliciting difficulty accounts relating to 'difficult' and 'not so difficult' patients (see 5.2.1.1.2 below) led to a new variable ('patient category') which could either serve as an independent variable (for

'difficulty type'); or as an alternative dependent variable, yielding information in its own right. In the event, both these options were followed up.

5.1.1 Links to the two previous studies

In the initial planning of investigations for this thesis, the main aim of a second survey had been to generate data relating to research question 3 (see 1.3 above), about the associations of difficulty types with measures of skills deficits and therapists' personal characteristics. This rationale still prevailed, but other issues also needed attention. The first concerned the revised difficulty types - the replication of their reliability with a different set of external judges, the confirmation of their relationship to each other, and the investigation of their correlates. The second pertained to revising the Familiarity, Impact and Duration scales - the familiarity measure was thought to behave in different ways depending on the response format, and the duration index seemed out of keeping with the other two scales. The third concerned the German-language versions of the FID scales, difficulty prompts and therapist variables - while these had been piloted, their psychometric properties and/or equivalence to the English-language versions had not been established.

The main focus of survey 2 was, however, the study of the relationships between transient and paradigmatic difficulty types and skills deficits or enduring personal characteristics of the therapists reporting them. The primary effort in constructing a new questionnaire was therefore directed towards this goal. Within this domain, paradigmatic difficulties had emerged as a central interest; partly because they extended previous theoretical understandings, partly because of the challenge they posed in the measurement of their correlates, and partly because of their links to countertransference and hence to the supervision of psychodynamic psychotherapists.

Some pointers of how the situational difficulty type might relate to measures of therapists' skills and personal attributes could be gleaned from TDQ1, however, no particular theoretical predictions were made at the outset of the second survey. In this respect (and in respect to the categories of 'difficult' and 'not so difficult' patients, see 5.2.1.1.2 below) the third study was exploratory and had to rely on a replication design for the confirmation of results. In other respects, however, general hypotheses were formulated and later operationalised for specific measures.

5.1.2 Hypotheses

There were three general hypotheses derived from research questions 2 and 3 (see 1.3 above). The first concerned the FID scales in their revised form used in TDQ2:

- Indices of the familiarity, depth of emotional impact, and duration of difficulties will be positively correlated with paradigmatic mean salience scores, but negatively correlated with transient mean salience scores.

The second general hypothesis concerned the link of transient difficulties to measures of therapists' competencies:

- Indices of therapists' lack of skills, knowledge, or experience in relation to a difficulty will be positively correlated with transient mean salience scores but not with paradigmatic mean salience scores for that difficulty.

The third general hypothesis concerned the link of paradigmatic difficulties to measures of therapists' enduring personal characteristics:

- Indices of therapists' internal conflicts in relation to a difficulty, or of the similarity of their personal characteristics to that of the patient with whom they experience the difficulty, will be positively correlated with paradigmatic mean salience scores but not with transient mean salience scores for that difficulty.

There was also one specific hypothesis about the link of transient difficulty type to the length of practical experience of the therapist reporting the difficulty. As one would have expected, a negative association between transient mean salience scores and practice length had been observed in TDQ1, but may have well been attributable to raters' knowledge of therapists' level of experience at the time of relating the difficulty account. This contextual information was omitted in the ratings of accounts collected with TDQ2 in order to test the hypothesis that:

- Transient mean salience scores will be negatively correlated with an index of the length of time a therapist had been practising when the difficulty first occurred.

5.1.3 English- and German-Language Versions

The instrument with which data were collected for this survey - the Therapist Difficulty Questionnaire Version 2 (TDQ2) - was simultaneously constructed in a German- and an English-language version[16]. The reasons for this strategy were one part design - two data sets from different languages provide enhanced opportunities for replication; one part pragmatism - if data collection in one language community should prove problematic there would be a fall-back position; and one part opportunism - it does not require too much additional effort but might turn up some interesting cross-cultural comparisons. If one of the language-samples was to be very small, it might still serve as a pilot for explorations to be confirmed by the larger sample. No predictions were made as to which of the two questionnaire versions was likely to provide the larger data set.

[16] In the event, the German version was printed first and the first data collected came from German psychotherapists (see 5.2.2.1.1.1 below).

5.2 Method

Initially, consideration had been given to using an interview rather than a survey format for the third study. For the reasons already outlined in 3.2 above, that notion was eventually rejected. Some design ideas were, however, tested out in two tape recorded interviews which were followed by debriefing sessions[17]. Having settled on a survey format, the general approach adopted was a refinement of the method used for survey 1. It employed a questionnaire which prompted recall of difficulty situations in detailed written accounts, followed by a number of variables relating to the experience of the difficulty, the patients with whom it was experienced, the respondents, and the interactions between the latter two.

The survey questionnaire (TDQ2) was designed with data analysis in mind. The general data analytic strategy, both for the exploratory and the hypothesis testing parts of the survey, was to use a replication design, identical to that outlined in the validation study (see 4.4 above). Data were divided into two sets ('batches'), ensuring that every case in each batch was drawn from a different respondent (see 5.2.4.2 below). Attention was paid both to the magnitude of associations and to their reliabilities. For magnitude, a correlation was deemed to be 'noteworthy' if it was in the same direction and above .15 (in absolute terms) in both batches. For reliability, combined probabilities from both batches were calculated using the formula given by Winer (1971) (see footnote 13, p. 105) and adopting a significance level of $p < .05$, which was adjusted for multiple comparisons where appropriate. The simplest statistical procedures deemed adequate were used throughout; multiple regressions were only employed where several predictors could theoretically be expected to relate

---

[17] One interesting notion to emerge from these interviews, which seems well worth following up in a qualitative study, was that of 'parallel process' (Mattinson, 1975) between interpersonal features of the difficulty reported and transactions in the research interview.

to a response variable. Principal Component Analyses were used to test newly designed scales for unidimensionality.

5.2.1 Instrumentation

Apart from questions tapping descriptive therapist and patient variables, four groups of instruments were included in TDQ2:

a) Prompts eliciting difficulty accounts (see 5.2.1.1.2 below);

b) measures of experiential qualities of difficulties - the FID scales (see 5.2.1.1.3 below);

c) measures of deficits in therapists' competencies - later supplemented by items tapping 'universality' and 'specificity' of a difficulty (see 5.2.1.1.4 below); and

d) measures of therapists' enduring characteristics and their perceptions of their patients' enduring characteristics - the IIP-51 (see 5.2.1.1.5.5 below), SASB INTREX (see 5.2.1.1.6 below), and the CCRT-Grid (see 5.2.1.1.7.5 below).

Some of these measures had to be newly constructed. Their design is described under the relevant sub-headings in the following section. The German version of TDQ2 used the already translated and validated versions of the IIP and SASB INTREX. All other measures were specifically translated.

5.2.1.1 Questionnaire Construction (English-Language Version)

Even though the German version of TDQ2 was the first to be printed and used for data collection (see 5.2.2.1.1 below), most of the development work was done in the English-language version which is therefore reported first. The different sections are described in the order in which they occur in the questionnaire.

5.2.1.1.1 Rationale

Following the experiences with TDQ1, there was reason to have confidence that therapists can call difficulty situations and related experiential states to mind when prompted to do so. Some of the accounts provided for the first study had been extremely brief and since it was clear that narratives of difficulty experiences collected with TDQ2 would again form the basis for salience ratings of difficulty types, efforts were made to obtain more extensive accounts in the hope that this would improve reliabilities.

It was crucial for the replication design of the study that respondents would again provide two accounts of difficulties and related measures each. The previous format of eliciting 'past' vs. 'current or very recent' difficulties was abandoned because of the disadvantages of having to design alternative versions of some measures and the associated problems with the Familiarity Scale (see 3.3.4 above). Instead, respondents were asked to provide one difficulty with a 'difficult' patient and another with a 'not so difficult' patient. Evidence that therapists can indeed distinguish between the two patient categories and that these differentiations are capable of external validation had been provided by Freedman (1992b) and Berzofsky et al. (1992) (see 2.1.2 above). Although the distinction had originally been made in the service of helping respondents recall two separate difficulties and providing a richer, more varied data base, it did yield, somewhat serendipitously, interesting information in its own right.

5.2.1.1.2 Eliciting Accounts

Narratives of difficulties with either patient category were elicited in an equivalent format. Respondents were asked to call to mind a situation which they had personally encountered in their practice of individual psychotherapy, which they had found

difficult, and which involved a patient whom they generally found difficult/not-so-difficult. They were then requested to write an account of the difficulty in some detail as it came to mind and, as far as possible, without censoring it. In order to facilitate a free-flowing narrative, the space provided for the account was not structured by sub-headings (as it had been in TDQ1). Instead, these were mentioned in the prompt as aspects of the narrative which might helpfully be included. Respondents were also invited to continue their account on the reverse of the page provided for their answer.

A section of the cover page (see 5.2.1.1.9.2 below) also prepared respondents directly for the task of providing accounts by suggesting ways in which they might focus on the experience (see appendix 8.1). Half the questionnaires distributed started with the 'not so difficult patient' account, the other half with the 'difficult patient' account, to correct for possible order effects.

5.2.1.1.3 FID Scales

As in TDQ1, the measures of Familiarity, Depth of Emotional Impact and Duration were placed immediately after the difficulty accounts under the heading "ABOUT YOUR EXPERIENCE OF THE DIFFICULTY". Only the 'past' versions of the scales were used, as the prompt for 'current' difficulties had been eliminated from the new questionnaire. Given the overall length of TDQ2 it seemed important to restrict the FID measures to a single page. This made it necessary to shorten the Impact and Familiarity Scales substantially; especially so, since space was also required to accommodate a new scale as described below.

5.2.1.1.3.1 Construction of a New Scale: Subjective Duration

The calculation of a Duration Index from the answers to questions about the 'objective' length of time for which a difficulty persisted, even though it yielded interpretable

results in TDQ1, contrasted sharply with the subjective measures of Familiarity and Impact. While therapists' estimates of objective duration might themselves be rendered inaccurate by subjective distortions, it nevertheless seemed more in keeping with the focus of the enquiry to construct another scale in order to tap therapists' subjective experience of duration[18]. Whether the two were actually distinguishable then became an empirical question which could be answered by including both measures in TDQ2, where the shortening of the Familiarity and Impact Scales had created space for additional items.

On reviewing likely conceptual associates of the dimension of 'subjective time', using Roget's Thesaurus, a distinction between words relating to 'absolute' time and those relating to 'relative' time emerged. This difference helped to elucidate further a facet of the duration concept which had so far remained ambiguous: notions associated with relative time such as 'recurrence' and 'newness' are close to concepts such as 'acquaintance' and 'novelty' which, being constituents of the Familiarity Scale, had hitherto been troublesome as they seemed to confound that scale with a time dimension. Using only connotations of absolute time for describing therapists' experiences of duration makes for better separation between dimensions and thus for greater conceptual clarity.

Three subordinate constructs of subjective experience of absolute time could be distinguished as follows.

---

[18] The distinction between two modes of experiencing time - 'chronos' (linear and objective) and 'kairos' (cyclical and subjective) - made, amongst others, by Modell (1990), contributed to the decision to investigate subjective duration.

'duration' (the generic category), denoting the spectrum of 'short' to 'long' duration;

'flux' (the idea of time as a stream), encompassing the construct of 'motion' vs. 'stagnation'; and

'permanence', including the construct of 'instanteneity' vs. 'perpetuity'.

Following the strategy adopted in the construction of the 'Familiarity' and 'Impact' scales, the dimensions described above were each represented by one item pair. Each pair consisted of a positively poled item, indicating long duration, and a negatively poled item denoting brief duration. The three item pairs which together form the new 'Subjective Duration' scale are shown in Table 5:1 below.

Dimension:	Polarity:	Item:
Duration	+	This felt like a long drawn out difficulty
	-	This felt like a brief difficulty
Flux	+	Time seemed to stand still during this difficulty
	-	This difficulty unfolded quickly
Permanence	+	This difficulty seemed to last an eternity
	-	This difficulty seemed to be over in an instant

Table 5:1- Subjective Duration Items

5.2.1.1.3.2 Familiarity Scale - Short Form

The 16-item Past Familiarity Scale in TDQ1 was reduced by successively removing items which made the smallest contributions to overall scale reliability until only ten items were left. These are shown in Table 5:2 below.

Dimension	Polarity	Item
Novelty	+	This kind of difficulty was well-known to me.
	-	Difficulties of this nature are new to me.
Frequency	+	I had experienced this sort of difficulty time and again.
	-	This was virtually the first time that I had encountered a difficulty of this type.
Recency	+	I had been acquainted with this sort of difficulty for a long time.
	-	It was then only recently that I had first met this kind of difficulty.
Uniqueness	+	I was used to this kind of difficulty.
	-	This was an unusual difficulty for me.
Acquaintance	+	I knew about difficulties of this kind through previous experience.
	-	I was not personally conversant with this type of difficulty.

Table 5:2- Shortened Familiarity Scale: Items grouped by Dimension and Polarity

Three dimensions - Anticipation, Surprise, and Typicality - were eliminated altogether in the shortening of the scale, leaving each of the remaining five dimensions represented by an item pair. The estimated reliability of the short scale was $\alpha = .93$ which is actually marginally higher than that for the long form. Subsequent analyses (see 5.3.3 below) revealed this estimate to be accurate.

5.2.1.1.3.3 Impact Scale - short version

The 24-item Past Impact Scale in TDQ1 was also reduced by successively removing items which made the smallest contributions to overall scale reliability. A target of ten items in the final version was aimed for, to equalise the length of both scales. The items constituting the short form are shown in Table 5:3 below.

Dimension	Polarity	Item
Impact	+	I was emotionally affected by this difficulty.
Impression	+	I experienced this difficulty as 'going to the core' of me.
Immersion	+	I was engrossed by this particular difficulty.
Contemplation	+	This difficulty gave me pause for thought.
Preoccupation	+	I was preoccupied by the experience of this difficulty outside the therapy session(s).
Penetration	+	This difficulty really 'got inside' me.
Effort	+	Grappling with this difficulty was a struggle.
	-	I found this difficulty easy to deal with.
Mindfulness	+	I felt concerned about this difficulty.
Dissociation	+	I could not let go of this difficulty.

Table 5:3 - Shortened Impact Scale: Items grouped by Dimension and Polarity

Only three dimensions were eliminated altogether - Concern, Control (which carried the notion of intrusive thoughts and images), and Saturation. On the other hand, only one dimension - Effort - survived intact. Of the other eight dimensions only the items with positive polarity were retained, raising the question whether these dimensions actually did represent bi-polar constructs. The estimated reliability of the short scale was $\alpha = .91$. Subsequent analyses (see 5.3.3) revealed this estimate to be accurate.

5.2.1.1.3.4 Combined Scale - Polarities and Ordering

The order of items from the long form of the combined scale was preserved in the short version. Subjective Duration items were interspersed in the gaps left by the elimination of items. Positive and negative versions of each of the three item pairs were placed equidistant from the middle of the scale to control for order effects. The ordering of items in the combined scale, together with their polarities and the concepts and dimensions they represent is shown in Table 5:4 below.

The visual appearance of the scale was changed by placing verbal anchors for scale points at the top of columns above the pie-chart icons and by separating items from each other by dotted lines (see appendix 8).

Polarity:	Scale:	Dimension:	Item:
+	IMP	Impact	I was emotionally affected by this...
-	FAM	Novelty	Difficulties of this nature were new..
+	IMP	Impression	I experienced this difficulty as...
-	DUR	Flux	This difficulty unfolded quickly.
+	FAM	Frequency	I had experienced this sort of...
+	IMP	Contemplation	This difficulty gave me pause for...
+	DUR	Permanence	This difficulty seemed to last an...
+	FAM	Recency	I had been acquainted with this sort of..
+	IMP	Penetration	This difficulty really 'got inside' me.
-	FAM	Uniqueness	This was an unusual difficulty for me.
+	DUR	Duration	This felt like a long drawn out...
-	IMP	Effort	I found this difficulty easy to deal with.
+	FAM	Acquaintance	I knew about difficulties of this kind...
+	IMP	Effort	Grappling with this difficulty was a...
-	FAM	Acquaintance	I was not personally conversant with...
+	IMP	Dissociation	I could not let go of this difficulty.
-	DUR	Duration	This felt like a brief difficulty.
+	IMP	Mindfulness	I felt concerned about this difficulty.
+	FAM	Uniqueness	I was used to this kind of difficulty.
+	DUR	Flux	Time seemed to stand still during this...
-	FAM	Recency	It was then only recently that I first...
+	IMP	Preoccupation	I was preoccupied by this difficulty...
-	FAM	Frequency	This was virtually the first time that..
+	IMP	Immersion	I was engrossed by this difficulty.
-	DUR	Permanence	This difficulty seemed to be over in...
+	FAM	Novelty	This kind of difficulty was well-...

Table 5:4 - Ordering and Polarities of Items in Combined Scale

5.2.1.1.4 Competence, Deficit, and Universality Items

In the original conception of a second survey, the construction of a measure of deficiencies in therapists' skills, knowledge and experience had been thought of as an important counterpart to measures of enduring therapist characteristics. After the revision of the research plan described in 4.2 above, and the inception of the category of 'situational difficulties', these items could also serve an additional function. In conjunction with two further items tapping 'universality' (pervasiveness across therapists) and 'specificity' (pervasiveness within therapists) of experiences of difficulty, they could now be used to let respondents classify their own difficulties according to the same criteria that external judges would use, but without being aware of the difficulty types or their definitions.

Self-reported competence deficits can be thought of either as a felt absence of a necessary quality or as the perceived lack of a resource. The first of these options was tapped by the question "How useful would the possession of each of the following have been in preventing you from experiencing the situation as difficult?", offering the choices of "a) more extensive theoretical knowledge", "b) a broader repertoire of technical skills", "c) a wider range of practical experience", together with the open category "d) something else (please describe)". Resource deficiencies were tapped by the question "How helpful would each of the following have been in equipping you to deal with the difficult situation?", offering the options "a) getting formal supervision or consultation", "b) reading relevant literature", "c) attending workshops or seminars", "d) being familiar with a wider range of clients", again supplemented by an open category "e) something else (please describe)".

Apart from skills deficits, which should be a marker for transient difficulties, the main distinguishing characteristic contained in the difficulty definitions (elaborated in 4.2.1) concerns the universality/specificity of the difficulty for the therapist describing it. Situational difficulties should be universally experienced as difficult, regardless of therapists' level of competence, while transient difficulties should be experienced as such by other therapists with a similar level of competence. Universality was assayed with the question: "How **difficult** would another therapist, of a theoretical orientation and a level of experience similar to yours, have found the situation?". Paradigmatic difficulties, in contrast, should be idiosyncratic to the therapist concerned and indicative of enduring personal characteristics. Consequently, one might expect similar difficulties to have occurred in respondents' other relationships outside therapy. Specificity was therefore assayed with the question: "How **often** have you experienced a difficulty of this kind in a situation outside therapy (that is in a family, social or work relationship)?

Answers to all questions and options consisted of ratings on a six-point ratio scale ranging from "0 = Not at all" to "5 = Very" without anchorings for the intermediate scale points, but with points visually displayed as equidistant from each other.

If therapists were able to classify their difficulties accurately according to the theoretical criteria, one would expect

a) transient salience to be positively associated with competence deficits and universality but negatively associated with specificity;
b) situational salience to be positively associated with universality but negatively associated with competence deficits and specificity; and
c) paradigmatic salience to be positively associated with specificity but negatively associated with competence deficits and universality.

5.2.1.1.5 The Inventory of Interpersonal Problems

The Inventory of Interpersonal Problems (IIP) (Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988) is a measure of interpersonal difficulties which people may encounter in their relationships with others. It forms one of the key components of TDQ2 as it is presented in three different implementations - therapists' ratings of their 'difficult' patients, therapists' ratings of their 'not so difficult' patients and therapists' self-ratings - and therefore allows for direct comparison between the patterns of respondents and those which they ascribe to the patients with whom they have experienced difficulties. Indices of similarity or dissimilarity can thus be computed and their associations with the difficulty types investigated.

5.2.1.1.5.1 Development of the Measure

The IIP was developed out of an investigation into the presenting complaints reported by patients entering psychotherapy which distinguished interpersonal complaints from symptoms and from disturbing cognitions about the self (Horowitz, 1979). Subsequently, a pool of such relationship problems was assembled, forming the basis for the development of an inventory (Horowitz, Weckler, & Doren, 1983) which, when tested on a student population, showed high internal consistency and promising test-retest reliability (Horowitz, French, Gani, & Lapid, 1980). Further revision (Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1987) resulted in a 127-item checklist which grouped difficulties in terms of either deficits (78 items starting with the phrase "It is hard for me to...) or excesses (49 items under the heading "...things that I do too much). An example of the former would be "It is hard for me to disagree with other people"; an example of the latter would be "I lose my temper too easily". Respondents rate on a five-point scale (ranging from 0 = 'not at all' to 4 = 'extremely') how distressing each of the problems described has been for them. In its

original form, the IIP yields a general index of severity of interpersonal problems as well as subscale scores which are described in more detail below.

Since its publication the IIP has gained acceptance as a supplement and an alternative to symptom-based checklists (such as the Brief Symptom Inventory (BSI) (Derogatis, 1975)). In Britain it has been included in the National Foundation for Educational Research assessment portfolio (Milne, 1990) and is under consideration for inclusion in a core battery of psychotherapy outcome measures (Barkham, Hardy, & Startup, 1996). It has been translated into other languages, including Italian, where discrepancies in subscale structure have been found, possibly due to characteristics of the sample used (Clementel Jones, Azzone, Battacchi, & Freni, 1996); Swedish, where subscales could be replicated (Weinryb, Gustavsson, Hellstrom, & Anderson, 1996); and German, where clinical validity was demonstrated by Schauenburg, Pekrun, & Leibing (1995). The German translation by Strauss (1994) was used in the German version of TDQ2 (see 5.2.1.2 below).

5.2.1.1.5.2 Circumplex vs. Factor Scoring

Two different strategies have been pursued in further investigations of the structure of the IIP. The first seeks to locate IIP domains within a circumplex structure, relating them to Leary's Interpersonal Circle (Horowitz, 1996), the second to refine subscales based on factor analyses.

5.2.1.1.5.2.1 Circumplex Subscales

The notion that IIP item scores might conform to the structure of the interpersonal circumplex had already been adumbrated by Horowitz (1979) and Horowitz et al. (1988). They noted that scores on both the IIP and the inventory from which it was derived could be more simply described by two higher order factors, corresponding

closely to the dominance and affiliation (nurturance) dimensions of the interpersonal circle. Alden, Wiggins, & Pincus (1990) further investigated the idea that the item set of the IIP "might be interpretable, within the context of interpersonal theory, as a 'problems' version of the interpersonal circumplex" (p.523). Following Horowitz et al. (1988) they used 'ipsatised' item scores, calculated by subtracting an individual's mean score from each single-item score, in order to remove a 'general distress factor' from the values and thereby increase differentiation between subscales. (See 5.2.1.1.5.4 below for a discussion of the advantages and disadvantages of this procedure.) Circumplex analyses were conducted on the item scores obtained from two independent non-clinical student samples, extracting two principal components and locating each item in the two-dimensional space defined by these components. After establishing the stability of the ordering of items across both samples, a subset of 64 items, defining eight separate octants (with theoretical midpoints at 45° intervals), was selected on empirical and thematic criteria and validated on a larger sample. Further validation of the circumplex ordering of the 64 items was obtained from the same sample by comparison with an established circumplex measure - the Interpersonal Adjective Scale (Wiggins, Trapnell, & Phillips, 1988).

The eight octants obtained in these studies, each of which represents a subscale of the IIP, were labelled (starting from the top and proceeding in counterclockwise direction) as:

Domineering - indicating problems arising from the wish to control, manipulate or change others;

Vindictive - implying problems in trusting others or being able to care about their needs or happiness;

Cold - indicating problems in feeling or expressing affection for others or in being able to forgive or make long-term commitments to them;

Socially Avoidant - denoting difficulties in initiating or maintaining social contacts and feelings of anxiety and embarrassment in the presence of others;

Nonassertive - suggesting problems in making needs known to others and in being firm and assertive;

Exploitable - pointing to problems in feeling or expressing anger towards others for fear of offending them, and to gullibility and vulnerability to exploitation;

Overly Nurturant - implying difficulties arising from trying too hard to please others and being too generous, trusting, caring or permissive;

Intrusive - indicating problems resulting from seeking attention or being inappropriately self-disclosing.

5.2.1.1.5.2.2 Factor Subscales

After construction of the original 127-item IIP, 12 subscales were defined and consensually labelled by independent raters (Horowitz et al., 1987). Like the items, subscales were grouped into those describing deficits (Hard to... - abbreviated as H.) and those expounding excesses (Too... - abbreviated as T.) as follows:

H. Assertive, Independent, **Intimate,** Self-Worth, **Sociable, Submissive;** and

T. Aggressive, **Controlling,** Dependent, Eager to Please, Hypersensitive, **Responsible**

In a preliminary study, mean scores on nine of these scales were shown to be related to the interpersonal circumplex, while three (H. Self-Worth, H. Independent and T. Dependent) were not accounted for by the affiliation and dominance dimensions. However, a factor analysis extracted only seven of these subscales empirically, and one of these did not prove to be stable in a replication study. The final six subscales are those shown in bold type in the list above, however, they incorporate only 83 out of the total 127 items. This solution therefore shows a certain lack of conceptual coherence, an issue which, together with some methodological shortcomings, has been addressed in the revision by Barkham, Hardy, & Startup (1994) described below.

5.2.1.1.5.3 Short Versions

In its full 127-item version the IIP is a lengthy and somewhat cumbersome scale, especially when used in conjunction with other measures for research or clinical screening. Attempts have therefore been made to shorten the scale while preserving the richness of information contained in the subscales. As described above, the version developed by Alden et al. (1990) for circumplex scoring halved the number of items, retaining 64. This version is also referred to as the IIP-C, C denoting circumplex.

Two further abbreviations of the scale have been published, one based on eight factor subscales (Barkham et al., 1996), drawing items from the full version, the other based on octants of a circumplex structure (Soldz, Budman, Demby, & Merry, 1995), drawing items from the 64-item version. Both halve the number of items again to 32, retaining four items for each of their eight subscales.

Barkham et al. (1994) criticise the factor subscales proposed by Horowitz on three methodological counts: a) *undersampling* - the factor analysis was based on fewer observations than variables; b) *sample bias* - women accounted for 86% of the sample; and c) *underextraction of factors* - using an arbitrary criterion of eigenvalues > 3, leading to too few (and thereby composite) factors being used in the solution. In their own study, the authors used a larger sample (N=250, giving a variable-to-observation ratio of 2:1) which was balanced for sex. Using an iterative procedure, extracting increasing numbers of factors within limits determined by an automated scree test, they arrived at an eight-factor solution as the highest stable and interpretable configuration. The resulting eight subscales, which have high reliabilities with alphas ranging from .80 to .96, are labelled "Hard to be Assertive", "Hard to be Sociable", "Hard to be Supportive", "Too Caring", "Too Dependent", "Too Aggressive", "Hard to be Involved" and "Too Open". Only the first two of these correspond to the factor scales proposed by Horowitz et al. (1988). In a further step the authors carried out another factor analysis in order to achieve an additional reduction of complexity. Following the strategy adopted by Alden et al. (1990), they used ipsatised scores - subtracting each subject's mean scores for all items from their individual item scores - in order to correct for a general 'complaints' factor. Four bipolar factors were found, denoting problems relating to competition (H. Assertive - T. Aggressive), socializing (H. Sociable - T. Open), nurturance (H. Supportive - T. Caring), and independence (H. Involved - T. Dependent). These appear to be conceptually closely related to the four psychosocial competencies proposed by Gilbert (1989), namely Competition, Co-operation, Care-giving and Care-eliciting.

Soldz et al. (1995) argue that the 64-item version of the IIP is conceptually more concise than the original full scale, as it was derived by choosing items that

maximised circumplex ordering. Using the IIP-C with a clinical sample, they selected items which correlated most highly with subscale means and validated the resultant 32-item version (referred to as the IIP-SC, SC denoting short circumplex) on two clinical samples, including one of patients diagnosed as personality disordered. Subscale alphas across all three samples ranged from .68 to .84 and total scale alphas from .88 to .89. Scale intercorrelations demonstrated that the circumplex ordering of items had been preserved.

As a result of the revisions described above, three reliable short versions of the IIP have now been published, two of which reduce the original scale by 75%.

5.2.1.1.5.4 Ipsatisation Issues

The strategy of removing a general factor from a scale by ipsatising scores has not been without controversy. Advocates of the procedure (Wiggins, Steiger, & Gaelick, 1981; Jackson & Helmes, 1979) contend that a self-report general factor represents systematic response bias, which is unrelated to item content and therefore decreases the validity of an instrument. Critics of the procedure (McCrae & Costa, 1983; Borkenau & Amelang, 1985) argue that a general factor may well have substantive meaning, which needs to be understood rather than eliminated in order to preserve the validity of an instrument. In terms of the IIP, the question is whether the general factor should be regarded as representing individuals' differential tendencies to endorse interpersonal problem descriptions and therefore be removed from scores; or whether it should be understood as representing a trait which may well reflect interpersonal difficulties and therefore be retained. An empirical answer to this question was sought by Tracey, Rounds, & Gurtman (1996) who investigated whether the IIP general factor was positively associated with other measures and whether

different levels of the IIP general factor would have a differential influence on scale scores as suggested previously by Gurtman (1992). They found the general factor to be positively related to scores on the BSI (Derogatis, 1975), as well as to measures of negative affectivity and self-deception in a non-clinical student sample. This may be seen as providing evidence for interpreting the general factor in a substantive way rather than in terms of response bias. However, a study with a different non-clinical sample did not show any influence of general factor levels on correlational dispersion, which may be interpreted as evidence that ipsatising does not decrease the validity of the IIP. In conclusion, it appears from their study that IIP scores can safely be ipsatised, but that the use of untransformed scores may yield additional information about respondents.

5.2.1.1.5.5 A Combined Short Scale - The IIP-51

An examination of the items used in each of the 32-item IIP versions described above showed a fair amount of overlap. The scales have 13 items in common and use a further 19 items each which are specific to the respective short forms. A combined scale, which allows for both factor and circumplex scoring (each based on a subset of 32 items), would therefore comprise 51 (13+19+19) items. As it is still appreciably shorter than the 64-item version, but retains twice as much information as each of the 32-item short forms, this composite 51-item IIP (subsequently referred to as the IIP-51) was assembled for inclusion in TDQ2. Table 5:5 below shows the combined scale and the assignment of each item to the relevant subscales - either in the set based on factors, or in the set based on circumplex octants, or in both. Items with membership of both sets are shown in **bold** typeface. The two items requiring reverse scoring are marked as '**Rev.**'. The sequence of items is congruent with the order in the original 127-item questionnaire.

	FACTOR	CIRCUM-PLEX
Hard to...		
1. **join in on groups**	H. Sociable	Soc. Avoidant
2. introduce myself to new people		Soc. Avoidant
3. confront people with problems that come up		Non-Assertive
4. **be assertive with another person**	H. Assertive	Non-Assertive
5. disagree with other people	H. Assertive	
6. **socialise with other people**	H. Sociable	Soc. Avoidant
7. **show affection to people**	H. Involved	Cold
8. feel comfortable around other people	H. Sociable	
9. understand another person's point of view		Domineering
10. **be firm when I need to be**	H. Assertive	Non-Assertive
11. **experience a feeling of love for another person**	H. Involved	Cold
12. **be supportive of another person's goals in life**	H. Supportive	Vindictive
13. feel close to other people		Cold
14. really care about other people's problems	H. Supportive	
15. put somebody else's needs before my own	H. Supportive	
16. take instructions from other people who have authority over me	H. Supportive	
17. feel good about another person's happiness		Vindictive
18. ask other people to get together socially with me		Soc. Avoidant
19. open up and tell my feelings to another person	T. Open **Rev.**	
20. **attend to my own welfare when somebody else is needy**	T. Caring	Over-Nurturant
21. be assertive without worrying about hurting the other person's feelings		Exploitable
22. tell a person to stop bothering me		Non-Assertive
23. let other people know when I am angry		Exploitable
24. keep things private from other people		Intrusive
25. make friends	H. Sociable	
26. make a long-term commitment to another person	H. Involved	
27. tell personal things to other people	T. Open **Rev.**	
28. be aggressive towards other people when the situation calls for it	H. Assertive	
29. be involved with another person without feeling trapped	H. Involved	
Too much...	**FACTOR**	**CIRCUM-PLEX**
30. I want people to admire me too much	T. Dependent	
31. **I open up to people too much**	T. Open	Intrusive
32. I am too aggressive towards other people		Domineering
33. I try to please other people too much		Over-Nurturant

34. I try to control other people too much		Domineering
35. **I put other people's needs before my own too much**	T. Caring	Over-Nurturant
36. I am overly generous to other people	T. Caring	
37. I lose my temper too easily	T. Aggressive	
38. **I tell personal things to other people too much**	T. Open	Intrusive
39. I keep other people at a distance too much		Cold
40. I let other people take advantage of me too much		Exploitable
41. **I am affected by another person's misery too much**	T. Caring	Over-Nurturant
42. I am too dependent on other people	T. Dependent	
43. I want to get revenge against people too much		Vindictive
44. **I argue with other people too much**	T. Aggressive	Domineering
45. I am too envious and jealous of other people	T. Dependent	
46. I am too suspicious of other people		Vindictive
47. I get irritated or annoyed too easily	T. Aggressive	
48. I am too easily persuaded by other people		Exploitable
49. I worry too much about other people's reactions to me	T. Dependent	
50. I want to be noticed too much		Intrusive
51. I fight with other people too much	T. Aggressive	

Table 5:5 - Items and Subscales of the IIP-51

Factor-based and circumplex derived subscales which correspond to each other are shown in Table 5:6 below together with their overlap in terms of the percentage of shared items:

Factor-based Subscale	Circumplex Subscale	Overlap Percentage
T. Caring	Over-Nurturant	75%
H. Assertive	Non-Assertive	50%
H. Involved	Cold	50%
T. Open	Intrusive	50%
H. Social	Socially Avoidant	25%
H. Supportive	Vindictive	25%
T. Aggressive	Domineering	25%
T. Dependent	Exploitable	0%

Table 5:6 - IIP-51: Overlap between Subscales

The lack of overlap between the T. Dependent and Exploitable subscales is congruent with the results obtained by Horowitz et al. (1987) (see 5.2.1.1.5.2.1) and appears to

support their contention that "subscales involving interdependence...seem to require more than two dimensions" (p.18). However, if one refers back to the interpersonal circle, e.g. (Crook, 1980), the description "docile-dependent" appears in approximately the same location as the 'Exploitable' octant, suggesting that both scales are indeed compatible.

5.2.1.1.6 SASB INTREX

In addition to the IIP, the other major research instrument employed in TDQ2 was SASB INTREX. Like the IIP, it is a published measure (Benjamin, 1983), has known psychometric properties (Benjamin, 1995), has been relatively widely used, is available in a shortened form (Benjamin, 1988a), and has a validated German translation (Tscheulin & Glossner, 1993). Unlike the IIP, it is also available in two parallel versions, and is supported by computer software, facilitating scoring and report generation for both clinical and research purposes (Benjamin, 1988b, 1997).

5.2.1.1.6.1 The SASB model revisited

The INTREX questionnaires are derived from the SASB model and coding system (Benjamin, 1974) which, as mentioned above (see 2.2.5.2), have been developed out of the Interpersonal Circle by expanding it into three different but interrelated versions ('surfaces') labelled 'transitive', 'intransitive', and 'introject'. Questionnaire items representing specific locations on the circumplex (one per octant in the Short Form, four to five items per octant in the Long Form) on each surface correspond to each other in accordance with the principles of complementarity outlined above. In contradistinction to the circumplex version of the IIP, SASB INTREX differentiates interpersonal behaviour into active and reactive components and adds an intrapsychic dimension. It further lends itself to the evaluation of specific situations and specific

states (for instance 'at best' or 'at worst'). The standard series of INTREX ratings - which consists of 'best' and 'worst' ratings for introject, relationship with significant other, relationship with both parents at age 5 - 10, and parental relationship - can easily be modified to suit particular areas of interest and has been adjusted in TDQ2 to cover specific difficulty situations.

5.2.1.1.6.2 Item Selection and Presentation

The version of the INTREX questionnaire used in TDQ2 (see appendix 8) is based on the 'Short Form' (Benjamin, 1995), which has one item for each cluster (octant) on each surface. Inspection of the items in one of the parallel versions, Version 1, showed some of them not to be readily transferable from their original North American to a British cultural context. An instance of this would be item 38: "X is recklessly neglectful of him or herself, sometimes completely 'spacing out'". The corresponding item in Version 2 reads "X carelessly lets go of him or herself, and often gets lost in an unrealistic dream world", which was thought to be more culturally appropriate. As Version 2 items generally seemed more acceptable, this version formed the basis for all implementations of SASB INTREX in TDQ2. However, in two instances Version 1 items were substituted as they appeared more appropriate. Given that both forms are held to be parallel, and that each item of the Short Form corresponds exactly to a defined point on the respective circumplex, these substitutions should not result in a loss of reliability. In three further instances minor modifications were made to Version 2 items, using the most recent revisions of the German items (N. Hartkamp, personal communication, April 1996) as a guide. For example, item 13 "With wonderful love and caring, X tenderly approaches if Y seems to want it" was changed to "With much love and caring..." to be congruent with the more attenuated emotional expressiveness characteristic of British (and, for that

matter, German) cultural norms. The possible minor loss of reliability which may have resulted from these changes was thought to be preferable to the loss of acceptability of the measure which might have resulted from the unmodified items.[19]

Instructions for the completion of SASB INTREX were taken from the original North American version with one minor idiomatic correction. The layout of the rating forms in TDQ2 is similar to the German Short Form (Tress, 1993) which appears better structured and visually less cluttered than Benjamin's printed response sheets.

5.2.1.1.6.3 Administration

There are three separate applications of SASB INTREX in TDQ2 - tapping respondents' private relationships, their transactions with difficult patients, and with not-so-difficult patients. Within the 'Therapist' ("About Yourself") section of the questionnaire, SASB INTREX Surface 3 ('Introject') ratings are asked for under the headings "How I treat myself in my private relationships **when I am at my best**" and "How I treat myself in my private relationships **when I am at my worst**" in order to obtain a baseline of therapist introject in non-clinical situations. Eliciting both states, following the recommendation by Benjamin (1995), has two purposes: it mitigates potential problems with social desirability response sets by giving respondents an explicit opportunity of showing themselves in the most favourable light, and it allows calculation of the distance between both states, which may provide an indirect index of intrapsychic pain / emotional disturbance (Benjamin, 1984; Davies-Osterkamp, Hartkamp, & Junkert, 1993).

---

[19] Despite these efforts, the INTREX measure was not universally acceptable. One respondent commented thus: "I thought the language was ... sometimes odd (tenderly and lovingly appreciate and value myself) - this is post-Thatcher England - not some neo-hippy co-counselling centre on U.S. west coast!"

The other two applications of SASB INTREX in TDQ2 are related to the accounts of difficulties with 'difficult' and 'not so difficult' patients. In both instances, respondents are invited, under the general heading "About what happened during the difficulty", to rate

a) their introject (Surface 3) in the course of the difficulty, prompted by the heading "How I treated myself during the difficulty";

b) their perceptions of their client's behaviour towards them, both active (Surface 1) and reactive (Surface 2), in the course of the difficulty, prompted by the heading "How my patient/client treated me during the difficulty"; and

c) their perceptions of their own behaviour towards their patient, again both active (Surface 1) and reactive (Surface 2), in the course of the difficulty, prompted by the heading "How you treated your patient/client during the difficulty".

Although the latter two difficulty-specific applications had been designated as 'optional' in order to provide respondents with an opportunity of shortening the questionnaire and thereby improve user-friendliness (see 5.2.1.1.9.1 below), more than 90% of respondents did, in the event, provide data for these sections.

5.2.1.1.7 The CCRT Grid

The last of the measures presented in TDQ2 attempted to assess the similarities between respondents and their patients in terms of their core conflicts, as measured by CCRT standard categories. For reasons outlined in 5.3.6 below, this attempt was unsuccessful and the measure proved to be confusing or unacceptable to many respondents. Its design is therefore reported here relatively briefly.

5.2.1.1.7.1 The CCRT model revisited

As described in 2.2.6.1.2 above, the 'core conflictual relationship theme' (CCRT) method had been originally designed as a means of identifying pervasive patterns in the way individuals conduct relationships and thus of transference (Luborsky, 1977). The basic structure of a CCRT - (unconscious) other-directed wish, need, or intention, followed by response-from-others (R-O) and response-from-self-to-response-from-others (R-S) - is reminiscent of the three-surface SASB structure though it does not map directly on to it[20]. As CCRTs are repetitive, enduring characteristics of individuals (Luborsky & Kächele, 1988), one would expect them to relate to paradigmatic difficulties, especially if a match between therapist and patient core conflicts could be established.

5.2.1.1.7.2 The Standard Categories

CCRT standard categories are a development from the 'tailor-made' categories (requiring judges to formulate individualised wishes and responses from relationship episodes) which are cumbersome and time-consuming (Luborsky, 1990b). Successive lists have been constructed by extracting the most frequently occurring categories from normative samples, first by Luborsky (1986) as cited in Luborsky (1990b), subsequently by Crits-Cristoph & Demorest (1988), and most recently by Barber, Crits-Cristoph, & Luborsky (1990). The latter was further developed into a 'cluster' version, comprising three list of eight composite categories for wishes, R-Os, and R-Ss respectively. This cluster version was used in constructing the CCRT grid in TDQ2.

---

[20] Comparisons between the two systems have, however, been made (Luborsky, Popp, & Barber, 1994).

5.2.1.1.7.3 Self-reported CCRTs: A previous attempt.

CCRTs, regardless of the category system used, are usually scored by external judges. One previous attempt to use self-report has been described by Kächele & Dahlbender (1993), who asked respondents to a questionnaire to rank-order each of the three lists of cluster-version standard categories in terms of their "general subjective significance"[21] for them. Categories of the same rank in each of the lists were combined into three-step CCRTs and respondents invited to reflect on the accuracy of such formulations. As the authors acknowledge, the degree to which CCRTs are conscious or subject to self-deception is moot. A different methodology was therefore adopted in TDQ2 for therapists' report of their own and (their perception of) their patients' CCRTs.

5.2.1.1.7.4 Repertory Grids

The use of repertory grids, following Kelly's (1955) role construct repertory test, has a long tradition in British clinical psychology (Winter, 1992). Respondents are asked to rate or rank a number of elements (typically significant others) in terms of a number of constructs, which are often individually elicited by presenting triads of elements, asking for ways in which two of these are alike but different from the third[22]. Although specifically elicited constructs are most in keeping with the conceptual base, predetermined ('supplied') constructs have also been used (Slater, 1976). A variety of software, some in the public domain, is available for the analyses of grids, usually based on principal component analyses and computation of angular distances between elements and constructs.

---

[21] "Allgemeine subjektive Bedeutsamkeit" (translation by the author).

[22] Interestingly, the proximity of the role construct repertory to the CCRT method was noted by Luborsky (1990a).

5.2.1.1.7.5 Combining CCRT and Grids

A major factor in deciding on the repertory grid as a format for measuring CCRTs lay in its capacity for measuring relationships between constructs and elements which may be outside a respondent's awareness.[23] Of the three elements of a CCRT, the last (R-S) should be most accessible to respondents' introspection. Conversely, the first (wish, need or intention) is most likely to be unconscious. It was therefore decided to use the eight 'wish' standard-categories from the cluster-version list as elements in the grid, each personified as "Someone who..." (for instance, "someone who wishes to assert self and be independent"). These elements were supplemented by the difficult and not-so-difficult patients, and two (conflictual) self-representations - "myself when I am most struggling in my role as a therapist" and "myself when I am most struggling in my private relationships".[24] This led to the construction of two grids (see appendix 8), one with R-Os, the other with R-Ss serving as constructs. It might be argued that it would have been more in keeping with the sequential nature of CCRTs to have used R-Os as elements in the second grid. Apart from this being potentially more confusing, it is also conceptually cogent to assume that respondents would have clear links between wishes and self-states in mind, based on their own experiences. This notion was expressed in the explanation accompanying the grid as follows (see appendix 8):

> "The **responses** in the **second** matrix consist of <u>possible reactions of the self</u>, which are **customary states** associated with the subject or evoked by the wishes/needs/ intentions described ('the subject becomes...')."

---

[23] Hicks & Nixon (1989) provide an example for a covert measure of self-concept by including self and selected others in the elements of a grid.

[24] The elements of the grid therefore had some resemblance to that used by Neimeyer, Klein, Gurman, & Greist (1983).

Explanations for the CCRT grid originally made reference to repertory grids. After piloting the measure with a German therapist, it became clear that this was not helpful as the methodology is not widely known in German-speaking countries. The explanations were therefore revised, expanded, supplemented with examples, and ran in the end to a whole page. Respondents were asked to form a mental picture of the person serving as the element (the 'subject'), and then to rate each construct, according to the likelihood with which it would be evoked by the subject, on a four-point scale ranging from '0 - not at all likely' to '3 - very likely'.

5.2.1.1.7.6 Indices generated by the CCRT Grid

The simplest measure to be derived from the instrument consists of the correlations between patient and therapist-self elements, calculated from ratings of constructs in both grids. These provide an index of similarity between therapists' perceptions of their patients and themselves. Further, if individual grids are analysed, their principal components form an n-dimensional space in which both elements and constructs can be located. Distances between relevant elements in that space potentially provide an index of therapist / patient dissimilarity.

5.2.1.1.8 The Feedback Section

As in TDQ1, the feedback section of this questionnaire consisted of two components - a three-item structured part and a free-format section inviting written comments on the questionnaire and the experience of completing it.

In the first part, respondents could endorse one of the options of 'too short', 'about the right length', or 'too long' in response to the prompt "Overall, I thought this questionnaire was..." for the first item, and 'easy', 'not too arduous', 'fairly laborious', or 'difficult' in response to a similar prompt for the second item. The third

item prompted "Completing this questionnaire felt...", giving the options 'rewarding', 'neither a bonus nor a burden', and 'an imposition'. The main purpose of all three items was to give respondents an opportunity of ventilating feelings after completing a taxing task rather than to elicit information. Scaling is therefore at best ordinal.

The open-ended part was also intended for self-debriefing of respondents but had, in addition, the potential to generate information about the acceptability of parts of the questionnaire and the validity of responses.

Finally, respondents were invited to contact the author if they wanted to discuss any aspect of the questionnaire or of their experience in completing it, giving them an opportunity of arranging debriefing, if they so wished.

5.2.1.1.9 Fostering a Research Alliance

In view of the length of this questionnaire and of the low return rates for TDQ1, particular attention was paid to developing and maintaining a 'research alliance'[25] with potential respondents. The three elements contributing to this effort were the initial letter (discussed below under 5.2.2.1), the covering letter and cover page, and measures taken to improve the user-friendliness of the questionnaire.

5.2.1.1.9.1 Improving user-friendliness

Recognising that TDQ2 is a long and somewhat unwieldy questionnaire, several devices were employed to facilitate its completion.

a) *Headings* such as "About your experience of the difficulty" or "About yourself" were placed at the top of sections, to help respondents orient themselves.

---

[25] Analogous to the components of the 'working alliance' outlined by Bordin (1979), one might think of the research alliance as consisting of an agreement on tasks and goals, and a bonding element, which has to be established and maintained, however, without the benefit of face-to-face contact.

b) *Colour coding* was used in an effort to break up the bulk of the questionnaire and to make it appear more manageable. The whole document was divided into four sections, each printed on different coloured paper, and respondents were invited in the cover page to counteract fatigue effects by completing one section at a time.

c) *Optional sections* were employed to give respondents some sense of control over their involvement. As in the TDQ1 conference sample questionnaire, two sections (SASB INTREX for transactions with patients) of the English-language TDQ2 were designated as 'optional', allowing respondents to reduce their total time commitment. In the event, only 6% of English-language respondents made use of this option, (compared with just under 10% of German-language respondents who omitted this section even though it had not been designated as optional) thus confirming previous experience of this strategy incurring little risk of data loss.

d) *Detachable instructions* were used with the CCRT-grid, which had the most involved set of directions. These were printed together on a single page so that they could be detached for easy reference without affecting the rest of the questionnaire.

5.2.1.1.9.2 The Covering Letter and Cover Page

The preliminary stages of involvement with the study were negotiated with most respondents through opt-in procedures and/or initial letters (see 5.2.2.1 below). Although their general interest could be presumed by the time they received the questionnaire, this next stage seemed crucial in building and maintaining their collaboration, now that it was plain what participation would involve.

With the truism that 'one never gets a second chance to make a first impression' in mind, each participant received in front of their questionnaire a personalised, hand-

signed letter on Trust notepaper, thanking them for their interest and their offer of help (see appendix 9). Potential benefits to themselves from their collaboration were elaborated to recipients, and they were prepared for the two different forms of responses (free-format and numerical ratings) contained in the questionnaire, requesting their patience with those parts which might seem of lesser relevance to them. After detailing return procedures and introducing the reply form, the letter finished with some information on the ISPDP (see 2.2.1.2.2 above), in case recipients were interested in that study, too.

The cover page (see appendix 8) was intended to deepen the alliance. It began by thanking respondents for 'embarking on completing this questionnaire', implying that by reading thus far, they had already enacted a commitment to carry through. Their reaction to the length of the questionnaire was then anticipated, validated and responded to by outlining reasons. Next, colour coding and optional sections were explained and offered as ways of making the task easier. Finally, respondents were prepared for the recall of difficulty situations required of them, followed by suggestions for ways of focusing on and subsequently disengaging from these tasks.

5.2.1.2 Questionnaire Construction (German-Language Version)

The German version of TDQ2 was constructed through translations of the English versions of its components by the author, with the exception of the IIP and SASB INTREX elements for which 'authorised' translations existed. These are reviewed below, together with those components for which translation was not entirely straightforward.

Although the long forms of the Familiarity and Impact scales used in TDQ1 had been translated into German, no empirical data were at hand as a basis for shortening them

prior to inclusion in TDQ2. English-language scale revisions, for which data were available (see 5.2.1.1.3.2 and 5.2.1.1.3.3 above), were therefore used in the construction of the German FID measures, together with the English version of the new Subjective Duration scale. As elements of some of the items were idiomatic (for instance 'going to the core' or 'got inside me'), translations for these had been of necessity less than literal. In one instance (item 18, "Diese Schwierigkeit wollte mir nicht aus dem Sinn") this may have led to problems (see 5.3.3.1 below). In general, however, translation of the FID measures did not appear particularly problematic.

A German version of the 64-item IIP-C has been published by Strauss (1994). Most of the items used in IIP-51 are included in the IIP-C, those which are not were translated by the author.

The INTREX long form questionnaire has been translated by Tscheulin & Glossner (1993) who confirmed circumplex ordering, construct validity, and content validity in a series of studies. Short form items are a subset of the long form questionnaire. They were selected to correspond with the English-language version (see 5.2.1.1.6.2 above).

The translations of CCRT standard categories as reported in Kächele & Dahlbender (1993) were used in the formulation of elements and constructs in the German CCRT grid. Some of these had to be slightly adapted to fit the format of 'personified' wishes and reactions (see 5.2.1.1.7.5 above).

5.2.2 Data Collection

Data collection with TDQ2 started in April 1996. The last returned questionnaire to be included in the data analysis was received in December 1997, although the bulk of data had been returned by the end of September of that year, about 18 months after

the start of collecting. The reasons for this extended period lie partly in the delays inherent in the means through which most data were obtained (opt-in mailings) which necessitated allowing respondents sufficient time for each stage, and in some cases a reminder letter. The main factor lengthening the process, however, was the administrative effort involved in contacting more than 1000 potential respondents directly, producing and distributing some 400 questionnaires, and compiling and managing the address lists on which the mailings were based.

TDQ2 is a long questionnaire. Pilots had shown it to require about 90 minutes to complete. In the event, extended difficulty narratives could lead to considerably more time being expended, as was evident from some of the comments made in the feedback section. As for TDQ1, the main concern was to recruit enough subjects into the study to allow for meaningful results. Although the mainly correlational design required attention to be paid as much to effect sizes as to probability levels, a larger sample would obviously give a better indication whether any weak associations were noteworthy. A power analysis (Cohen, 1988, 1992) was not carried out because there was no indication from previous research as to the magnitude of effect sizes which one might expect. It was thought, however, that a sample of 50 respondents from one language community, and of 25 from the other, would be a minimum to aim for. It soon became apparent that English language data were more readily obtainable, and data collection then concentrated on identifying and contacting successive groups of potential participants. In the end, a decision to discontinue collecting further English-language data when a target of 100 respondents had been reached, was governed by constraints on time and energy rather than by the conviction that an optimum number had been achieved.

Data collection with TDQ2 was not an impersonal exercise. The interactive nature of some of the sampling strategies and the number of potential respondents personally or professionally acquainted with me created the subjective impression of a dialogue with many participants. The last stage of this dialogue - the dissemination of results to respondents - is yet to come, but so far the study appears to me more characterised by collaboration than by detached observation. The resulting advantages and disadvantages are taken up in the discussion in chapter 6.

5.2.2.1 Sampling

As in the study carried out with TDQ1, 'representativeness' of the sample was not a goal of the data collection. The same conceptual reasoning - that the population from which samples could be drawn is of unclear delineation and unknown characteristics - applies, and similar practical constraints obtained - principally the expected low return rate. The main sampling strategy was therefore again one of identifying a target group which could be expected to be responsive to an appeal for co-operation, of trying to maximise returns by finding additional means of engaging respondents, and of selecting successive further target groups. Within this general strategy, efforts were made, however, to increase the richness of the data by trying to include groups with divers theoretical persuasions and professional backgrounds. To what degree these efforts succeeded is indicated by the characteristics of the composite sample discussed in 5.3.1 below.

The various samples are described below, grouped by language and data collection strategies. Return rates are summarised and discussed at the end of the section.

5.2.2.1.1 The German-Language Sample

The German version of TDQ2 had been completed before the English version and data collection began with German-speaking psychotherapists, although at that point it was not clear which language community would provide the main sample.

5.2.2.1.1.1 Direct Distribution: The LPW Conference Sample

The first data collection for the study was undertaken in the first week of the 46th Lindauer Psychotherapie Wochen (LPW). This is an annual two-week event providing one of the major sources of Continuing Professional Development for German-speaking psychotherapists. Following the strategy previously employed (see 3.2.2.1.1), a paper (Schröder, 1996) summarising previous work in relation to therapist difficulties and coping strategies, but keeping the audience unaware of the specific hypotheses currently under investigation, was presented as one of a series of invited lectures on psychotherapy research. Those present were invited to take a copy of TDQ2 and return it either during the conference or later on by post. A print run of 65 questionnaires was distributed on this occasion. A total of 12 questionnaires were returned during the conference and a further 2 arrived later by mail.

5.2.2.1.1.2 Postal Distribution: Swiss CCQ Respondents

Respondents to the ISPDP CCQ (see 2.2.1.2.2 above), who wish to extend their involvement in the study, can opt to become either 'corresponding' or 'active' members of the Collaborative Research Network (CRN). Doing so implies a willingness in principle to take part in further research projects connected with the ISPDP. The CRN steering committee agreed at its meeting in December 1995 to give TDQ2 the status of an ISPDP research module, thereby opening access to CRN members as a further sample for the current inquiry. While attending a CRN steering

committee in Bern in September 1996, an attempt was made to collect data from German-speaking Swiss CRN members. As it transpired, most Swiss CCQ respondents had not been supplied with a form giving them the opportunity of opting to become CRN members, leaving only a small sample of 16 corresponding members. These received a mailing of the questionnaire together with a covering letter explaining the connection of this research module with their previous involvement, a reply form and return envelopes.

No returns have been received from this sample. Possible reasons for this unexpected outcome may be found in the absence of any additional means of engaging respondents (who might have experienced the sudden receipt of a lengthy questionnaire as a form of 'cold calling') and in the fact that corresponding members had not received a promised bi-annual newsletter informing them about CRN activities. In the light of this experience, plans to contact German CRN members were held in abeyance pending the issue of a further newsletter, and it was decided not to contact CRN members from the UK without providing them with a newsletter and possibly a relevant paper as additional means of restoring and maintaining a research alliance. Neither of these two groups of potential participants was subsequently approached.

5.2.2.1.1.3 Opt-In Mailing: German-Language SPR Members

The strategy of inviting potential respondents by letter to participate in the research, and of sending them the questionnaire on receipt of a personalised reply postcard was adopted partly for practical reasons, as a means of saving on the costly production of questionnaires, and partly as a way of building a research alliance with participants by interacting with them and placing them in control of receiving the research materials.

Members of the international Society for Psychotherapy Research (SPR) might be expected to be particularly interested in and sympathetic towards a study conducted by someone from within their own ranks and addressing their allegiances both as psychotherapists and as researchers. A list of all SPR members was obtained via e-mail from the Society's executive officer. A total of 144 German members, 10 Austrian members and 28 members from the German-speaking part of Switzerland were identified and entered into a database (Microsoft Word Mailmerge). Letters (see appendix 9) sent to this group made reference to recipients' SPR membership, linked the work to previous collaborative research and stressed the bi-national aspect of the study. Six letters were returned as they could not be delivered, 6 recipients wrote back to decline, and 58 replies requesting questionnaires were received. About twelve weeks after despatch of the questionnaires, reminder letters (see appendix 9) were sent to those participants who had not returned a reply form by then. These included an anonymous reply postcard on which potential respondents who had decided against taking part were invited to give their reasons. Three therapists availed themselves of this option. The principal reason given for declining (after either mailing) was lack of recent therapeutic experience with individual clients. Apart from this, two recipients voiced doubts about the methodological basis of the study, two cited lack of spare time as explanation for opting out, one felt too involved with one of the instruments (SASB INTREX) used, and another felt the enquiry to be too intrusive into his 'private space' ('Intimsphäre'). The total number of completed questionnaires eventually received from this sample was 15.

5.2.2.1.2 The English-Language Sample

5.2.2.1.2.1 Questionnaire Distribution: Local Samples

The first English-speaking group targeted for TDQ2 consisted of local professional colleagues and fellow psychotherapists who could be expected to be accommodating or to feel a sense of personal obligation. Potential respondents received the questionnaire (together with a reply form and separate envelopes with different return addresses), accompanied by a personalised covering letter modified to apply to their particular circumstances (see appendix 9). The main subgroups in this sample consisted of :

a) 12 members of the Clinical Psychology Department within Southern Derbyshire Mental Health (NHS) Trust (SDMHT);

b) 14 members of the SDMHT Psychotherapy Department;

c) 8 members of the SDMHT Behaviour Therapy Unit;

d) 30 members of the NHS Psychotherapy Departments in Leicester, Lincoln and Nottingham, which, together with the Derby Department, are members of a collaborative specialist training scheme - The South Trent Training in Dynamic Psychotherapy.

A total of 64 questionnaires were distributed to this sample. Since questionnaires used in the British data collection did not carry sample codes, return rates are based on the receipt of completed reply forms. A total of 9 reply forms from local samples were received.

5.2.2.1.2.2 Opt-in Mailings

The rationale for this form of soliciting respondents has been outlined in 5.2.2.1.1.3 above. The two different approaches used in the British data collection were personalised and general opt-in mailshots as explained below.

5.2.2.1.2.2.1 Personalised Opt-in Mailings

These covered target groups for which lists of individual addresses were available, allowing for the generation of personalised letters referring to common interests or other links with the author. A higher opt-in rate was expected for this form of data collection as compared to general (not personalised) opt-in mailings.

5.2.2.1.2.2.1.1 British SPR Members

This sample was extracted from the SPR membership list referred to in 5.2.2.1.1.3 above. A total of 192 letters were sent to this target group. Those that were addressed to recipients known to me personally also carried hand-written personalised additions. Letters were timed to arrive in the week before the 1997 annual meeting of the SPR (UK) chapter, which I attended, to allow for personal contact and follow-up with those (approximately 60) members present at the conference. Eight letters could not be delivered and were returned. Eleven recipients wrote to decline participation (some after receiving the questionnaire). The main reason given was lack of current therapeutic involvement with individual clients; some replies also cited pressure of other commitments as a reason for opting out of the research. A total of 45 questionnaires were requested and sent out to this target group. As for the German SPR sample, reminder letters had been sent out to non-responders. Eventually, 21 reply forms from this sample were received.

5.2.2.1.2.2.1.2 Chartered Counselling Psychologists

The designation 'Chartered Counselling Psychologist' is a recent addition to the field of professional psychology, deriving from the establishment of a Division of Counselling Psychology within the British Psychological Society (BPS) in 1994 (Woolfe, 1996). As a new and still relatively small group, counselling psychologists could be expected to feel some commitment to supporting research efforts from someone in their ranks, in addition to having an investment in therapeutic practice based on a body of knowledge grounded in research. After being elected to the Division of Counselling Psychology early in 1997 as a full member and thereby being entitled to sign myself 'Chartered Counselling Psychologist', I wrote to all 200 Chartered Counselling Psychologists listed as such in the 1996 BPS Register of Chartered Psychologists, inviting them to opt in to the study. A small number of letters, whose addressees were known to me, also had hand-written personal messages appended. Five letters were returned undelivered. Ten recipients wrote to decline, mostly because of lack of current involvement in individual psychotherapy. A total of 44 questionnaires were sent out as requested, resulting in the return of 21 reply forms from this sample.

5.2.2.1.2.2.1.3 Selected Chartered Clinical Psychologists

The database for this group was assembled by going through the BPS Register of Chartered Psychologists and extracting the names of all members with whom I had previously come into contact, but who were not members of other target groups. A total of 26 letters were sent out, all of which had hand-written personalised messages appended to the text. Two recipients wrote back to decline: one was no longer engaged in clinical work; the other (who had been my tutor during basic professional

training) delivered himself of a heartfelt condemnation of quantitative research in this area in general, and the approach chosen by me in particular. Eleven questionnaires were sent out to this sample as requested, resulting in three completed reply forms.

5.2.2.1.2.2.1.4 Consecutive Chartered Clinical Psychologists

Although not all clinical psychologists are engaged in the delivery of psychological therapies, those that are might be expected to carry some allegiance to the scientist-practitioner model and to have a more positive attitude towards clinical research than some other professional groupings. Although providing a weaker basis for a research alliance compared with the other samples, this group constituted a large pool of potential respondents. Consecutive names of Chartered Clinical Psychologists were extracted from the BPS Register of Chartered Psychologists and assembled into a database provided they did not meet one of the following exclusion criteria: a) membership of another sample, b) exemption from needing a Practising Certificate (usually indicating clinically inactive members), and c) lack of indication from the entry that the member was involved in clinical work (absence of work address, address of non-clinical workplace etc.). The first 96 members (with surnames beginning with A, B, and C up to 'Campbell') received opt-in invitations. Further Chartered Clinical Psychologists might have been contacted had the data collection target not been reached. There seems no reason to assume that approaching consecutive rather than randomly selected recipients would have introduced a particular bias. One letter was not delivered and returned. One recipient wrote to decline because his NHS trust was being merged, leaving him with little spare time or energy. Another recipient wrote to enquire about the basis on which she had been targeted, but, on receiving a reply setting out the criteria, did return a reply form. A

total of 15 questionnaires were requested from this group resulting in 5 returned reply forms.

5.2.2.1.2.2.1.5 TDQ1 Respondents

A further target group for data collection consisted of previous respondents to TDQ1 as described in section 3.2.2 above. A total of 22 subjects in the study had indicated their willingness to collaborate with a further stage of the study and had supplied their names and addresses by returning a reply form separately from the main questionnaire; 21 of these were contacted, the remaining one being excluded as she was in training therapy with me at the time. All of the contacts received a copy of a conference paper (Schröder & Davis, 1994) which updated them on a previous stage of the study without making them aware of the hypotheses currently under investigation, together with a covering letter (see appendix 9) inviting their participation in the last stage of the study. One recipient declined participation by letter because he had moved from being a clinician to being a manager. A total of ten requests for questionnaires were received from this group, resulting in two returned reply forms.

5.2.2.1.2.2.1.6 University of Warwick MSc. Graduates

A total of 64 students had been enrolled for the MSc. in Psychotherapy at the University of Warwick during the 16 years of its existence. Most of these were personally known to me, either from the early stages when I was a student on the course, or from the last ten years when I was involved as a visiting speaker, group conductor or supervisor. Members of this constituency might be expected to be positively disposed towards the study, remembering their own struggles as psychotherapy researchers. On the other hand, unresolved left-over feelings from the

course might also be expected to get in the way of participation. Opt-in invitations were sent to 41 former graduates; all of them carried hand-written personalised messages. The letters were timed to coincide with a course reunion which provided opportunities for personal contact and follow-up. A total of 26 questionnaires were sent out on request, resulting in nine returned reply forms.

5.2.2.1.2.3 General Opt-in Mailings

The invitations to opt in to the study which fall under this heading were sent out individually to potential respondents but could not be personalised, either because the organisation concerned would not give access to its membership list or because the mode of distribution did not allow for separately addressed communications.

5.2.2.1.2.3.1 NAPP Members

The Nurses Association for Psychodynamic Psychotherapy (NAPP) is open to registered nurses who have undertaken a training in psychodynamic psychotherapy (full members) and to other nurses working in this and related fields with a general interest in the approach (associate and student members). A letter inviting potential participants to opt into the study by completing a pre-addressed reply postcard or by getting in touch via phone, fax or e-mail, was sent out with a conference mailing in June 1997. Although a total of 152 letters were sent out, about 60 of these would have been received by associate and student members. A small number of the remainder would have been sent to individuals previously canvassed as part of the SPR(UK) sample, and a further proportion would have come to recipients who are not currently engaged in psychotherapy with individuals. Therefore, the actual number of new potential respondents reached by this procedure might be estimated as approximately

75. Five questionnaires were requested by NAPP members resulting in three returned reply forms.

5.2.2.1.2.3.2 Trent Region Clinical Psychologists

Most clinical psychologists within the Trent Region receive 'Training Link', a bi-monthly newsletter jointly published by the professional training courses at Leicester and Sheffield Universities. A one-page insert (see appendix 13), setting out the main points of the study and inviting potential participants to opt in by completing and returning a tear-off slip or by getting in touch via phone, fax or e-mail, was sent out with the August 1997 edition of Training Link. This was timed to coincide with the publication of an article outlining my personal views on supervision which made explicit reference to therapist difficulties (see appendix 13). Although a total of 350 inserts were distributed, approximately 90 of these were sent to departments and institutions, leaving about 260 opt-in invitations to reach individual clinical psychologists directly. A number of these would have been sent to individuals previously canvassed as part of other samples, while another proportion would have come to recipients who are not engaged in psychological therapies with individuals. Therefore, the actual number of new potential respondents reached by this procedure is likely to have been closer to 200. This mailing resulted in 19 requests for questionnaires leading to 8 completed reply forms.

5.2.2.1.2.4 Advertisements

Two professional newsletters, covering psychotherapists and counsellors respectively, were chosen to carry advertisements requesting readers to consider taking part in the study. A third, addressing cognitive-behavioural therapists, was not pursued because of the poor response it had generated to a previous advertisement of mine, requesting

help with a different questionnaire. This avenue to approaching potential participants was not expected to generate many requests for questionnaires for the following reasons: a) the advertisements (see appendix 14) were limited in the amount of information they could convey, b) potential readers are likely to vary so widely that it is difficult to think of a good basis for forming a research alliance, and c) the newsletters carrying the advertisements are not always perused very carefully or promptly - a couple of replies were received several months after the publication dates with comments indicating that respondents had just noticed the invitation to participate.

5.2.2.1.2.4.1 UKCP Newsletter

This is a quarterly publication, distributed to psychotherapists registered with or affiliated to the United Kingdom Council for Psychotherapy (UKCP). Readership is potentially wide, with at least 2000 registrants receiving the newsletter. Ten requests for questionnaires were received from this source, resulting in the return of five completed reply forms.

5.2.2.1.2.4.2 CPCT Newsletter

This is a bi-monthly publication, distributed to subscribers to the Counselling in Primary Care Trust (CPCT). It has a print run of about 600 and reaches a readership engaged in or interested in professional counselling in GP practices. Three requests for questionnaires were received from this source, resulting in one completed reply form and one recipient declining by letter because of the length of the questionnaire.

5.2.2.1.2.5 Web-site posting

The plan to make the questionnaire available to potential respondents via the Internet was discussed and agreed with the person maintaining the SPR(UK) web-site. In the

event, the questionnaire was not mounted because of the anticipated technical difficulties for respondents in completing it on-line, or downloading and printing it out in the correct format.

5.2.2.1.3 Return Rates

Return rates for all German-language samples are summarised in the table below:

Sample:	Target Group Size	RTS	Decl.	No. of Questnrs. distribtd.	% of Target Group	Reply Forms retrnd.	% of Target Group	% of Questnrs. distribtd.
LPW	**65**			**65**	*100%*	**14**	*23%*	**23%**
Swiss CRN	**16**	1		**16**	*100%*	**0**	0%	**0%**
SPR Austria	**10**	1		**5**	*56%*	**2**	22%	**40%**
SPR Swiss-German	**28**	2	4	**8**	*31%*	**2**	8%	**25%**
SPR Germany	**144**	6	6	**45**	*33%*	**10**	7%	**22%**
Total	**263**	10	10	**139**	*55%*	**30**	12%	22%

Table 5:7 - TDQ2: Return Rates for the German-Language Sample

The first column of the table shows the number of potential participants either written to with an invitation to opt in to the study, or - in the case of the LPW conference sample - provided with the questionnaire directly. The second column gives the numbers of letters returned to sender (RTS) because the address was incorrect or the recipient deceased. Respondents who declined by letter are listed in the third column. The fourth column reports the numbers of questionnaires actually sent out or distributed; and the fifth column shows this figure as a percentage of the target group

across the samples, excluding those who could not be contacted (RTS). Column six reports the number of reply forms received and columns seven and eight show this figure as percentages of the target group and the questionnaire distribution respectively. The number of questionnaires returned was 30 although only 28 reply forms were received. Sample codes on the German-language questionnaires show that 15 questionnaires each were returned from the LPW sample and from the combined SPR samples. Figures in the total columns have been adjusted to take account of the two additional respondents.

Return rates are low by either measure used, especially if one takes into account that the Austrian response rate was inflated by the late return of a questionnaire which contained just over half the data asked for. With this proviso, there is a fair consistency in the German-language sample in terms of the figures in the last column. Between 20% and 25% of the questionnaires sent out or distributed were completed - with the exception of the Swiss CRN sample discussed in 5.2.2.1.1.2 above. Uptake rates between the larger Swiss and German SPR samples are also quite similar at about one third of the target groups.

Return rates for the English-language samples are shown in Table 5:8. The layout of columns follows the format of the previous table. No target group sizes are shown for the two advertisements as it is impossible to estimate how many readers of the two newsletters actually noticed the advertisements. Because of the uncertainty about the size of the target group in the two general opt-in mailings (NAPP-members and Trent Region Clinical Psychologists), distribution and return rates based on this group size are shown in brackets.

Sample:	Target Group Size	RTS	Decl.	No. of Questnrs. distrib.	% of Target Group	Reply Forms retrnd.	% of Target Group	% of Questnrs. distrib.
Local Samples	**64**			**64**	*100%*	**9**	14%	**14%**
SPR(UK)	**192**	8	11	**45**	*24%*	**21**	*11%*	**47%**
Counselg. Psychols.	**200**	5	10	**44**	*23%*	**21**	*11%*	**48%**
Selected Clinical Psychols.	**26**		2	**11**	*42%*	**3**	*12%*	**27%**
Consectv. Clinical Psychols.	**96**	1	1	**15**	*16%*	**5**	*5%*	**33%**
TDQ1 Respon-dents	**21**		1	**10**	*48%*	**2**	*10%*	**20%**
Warwick Grads.	**41**	1	2	**26**	*65%*	**9**	*23%*	**35%**
Total 1	**640**	15	27	**215**	*34%*	**70**	*11%*	**33%**
NAPP Members	approx **75**			**5**	*(7%)*	**3**	*(4%)*	**60%**
Regional Psychols.	approx **200**			**19**	*(10%)*	**8**	*(4%)*	**42%**
UKCP Advert.				**10**		**5**		**50%**
CPCT Advert.			1	**3**		**1**		**33%**
Total 2				**37**		**17**		**46%**

Table 5:8 - TDQ2: Return Rates for the English-Language Sample

The row labelled 'Total 1' gives summary figures for all samples for which a target group could be accurately determined. 'Total 2' summarises return rates for the other samples. As there were 17 fewer replies than completed questionnaires, return rates

are bound to be an underestimate. However, the absence of sample codes makes it impossible to apportion the additional returns to samples. Overall, 252 questionnaires were distributed and 104 (41.2% of these) were returned, almost twice the proportion of the German sample. However, expressed as a percentage of the target group, the aggregate figure (11%) is quite similar to the German sample. Not surprisingly, this percentage is lowest for the sample of consecutive Clinical Psychologists, which had the most tenuous basis for a research alliance of all the personalised opt-in samples. The highest percentage is found in the sample of Warwick graduates, which might be expected to have had the strongest basis for such an alliance. Another salient feature is the close similarity between the two large opt-in mailings, SPR(UK) members and Chartered Counselling Psychologists. Although they have a relatively low uptake (just under a quarter of all potential participants opted in), the return rates of questionnaires are high (just under half of those that had been sent out).

Figures from both samples confirm that return rates were very low. A great deal of effort had to be expended to secure the number of questionnaires eventually collected. The respondents are therefore likely to be a highly self-selected group whose characteristics are detailed in 5.3.1 below.

5.2.2.2 Ethical Considerations

The ethical issues raised by TDQ1 and TDQ2 are quite similar and are therefore considered here in respect of both studies. Following Korchin & Cowan (1982), three main areas of concern can be distinguished: informed consent, reduction of potential harm, and privacy and confidentiality.

Respondents were not asked to sign an informed consent form because it was assumed that they would have enacted their consent by returning the questionnaire -

in the same way that one might accept the terms of a software license by breaking a seal. Care was taken, however, to provide information about the studies both in the initial communications and in the cover pages of the questionnaires. Possible benefits and costs to participants were addressed in the latter (see 5.2.1.1.9.2). Potential respondents were also invited to contact the author, and were provided with a postal address, telephone and fax number, and e-mail address, in case they wanted to discuss the study before making a decision to participate. Care was taken, too, not to pressure participants, especially in the reminder letters, in order to ensure that consent could be given freely. Apart from research ethics, these measures also address concerns about the quality of the data collected, as they can be understood as ways of establishing and maintaining a research alliance.

Harm to respondents in psychological studies "is most likely to come from such things as stirring up painful feelings or memories, threats to one's self-image or embarrassment" (Barker, Pistrang, & Elliott, 1994, p.189). Clearly, all of these were potential risks to respondents to TDQ1 and TDQ2. Three respondents to TDQ2 did indeed comment about the disturbance caused to them by calling to mind difficult situations, although this is more than balanced by the number of respondents who commented that this had been a beneficial exercise. The possible emotional impact of completing the questionnaire was explicitly addressed in the cover page of TDQ2; and respondents were invited to disengage actively from their recalled state after completing a section of the questionnaire. The feedback section provided respondents with an opportunity to debrief themselves and a number of participants used the opportunity to discharge feelings of anger or frustration with parts of the questionnaire or with the study as a whole. Contact addresses and numbers were reiterated on the last page together with another invitation to discuss any aspects of

the study. Had respondents suffered distress from the completion of the questionnaire, they would have had a chance to arrange a personal debriefing but, in the event, no one took up this option. Finally, as all respondents were by definition active psychotherapists, one could reasonably assume that they would have provided themselves with means of support, consultation or supervision in respect of their work, including the work they were reporting in both questionnaires.

Protecting privacy by ensuring that participants in the studies could not be identified is again an issue of quality as well as one of ethics. As both questionnaires relied heavily on respondents' openness and willingness to expose potentially painful parts of private experience, any visible safeguard aimed at protecting anonymity was likely to increase the likelihood of honest replies. Two English respondents commented specifically on this issue: One spoke about struggling with her reluctance to expose what she regarded as a therapeutic blunder in an "ostensibly confidential questionnaire"; another expressed concern about his account of a very unusual difficulty being published and circulating among therapists "as a good trade-story". Separating questionnaire returns from response forms and providing different addresses for both should have gone some way towards reassuring respondents about anonymity. As far as intrusion into privacy is concerned, respondents were in control of how much to disclose. Codability indices in the reliability studies attest to the variability in degrees of disclosure, and participants were free to decline taking part, as one of the German potential respondents did, making explicit reference to the protection of his "Intimsphäre". Deciding how much of the difficulty accounts will become public is a more complicated question. Clearly, the very fact of incorporating difficulty accounts into a thesis highlights that, while anonymity can be safeguarded, confidentiality cannot be preserved (Sieber, 1992). To what degree access to the

source material should be constrained is more problematic. As far as access to this thesis is concerned, the utilitarian argument (Deyhle, Hess, & LeCompte, 1992) obtains - benefits accruing to the community of psychotherapists from studying difficult aspects of their practice might well outweigh individual costs as long as anonymity is ensured. Publication of difficulty accounts may, however, lead to a possible conflict of interests between the author and respondents who may prefer not to come across their descriptions of their struggles in a journal paper. Anonymity of replies makes direct negotiation with respondents impossible. Possible solutions to this dilemma would include the disguise of material and the observation of a lengthy interval before making accounts public.

Reciprocity between researchers and participants is of particular concern to qualitative researchers (Miles & Huberman, 1994) and has been a feature of the data collection with TDQ1 and TDQ2, as noted above. The final stage of that interaction - feedback of results to participants - will have to wait until completion of this thesis. The ethical responsibility there is not only to respondents but to the research community, as therapists are less likely to collaborate in future projects if promises to them from a previous study have been broken.

The study was not submitted for approval by an Ethical Committee. The use of a survey format, in which respondents were not identifiable, would generally give it 'Exempt' status (Barker et al., 1994) because it poses minimal risk. To ensure that this view accorded with local procedures, the rules of the Ethical Committee for the Southern Derbyshire Health Authority (which, at the time, was the body setting out the regulatory framework for the employment in which the research was conducted) were obtained and it was verified that there was no requirement to seek approval for this type of study.

As the study was completed before October 1998, when new data protection legislation is due to come into force, it is subject to the 1984 Data Protection Act. The computer used to generate mailings from address lists was registered in accordance with the act. Questionnaire data held on computer files cannot be connected with individuals and are therefore outside the scope of the act. Reply forms from TDQ1 and TDQ2, which do contain respondents' names and addresses, have only been kept as hard copies. Consequently, they are not covered by data protection legislation.

5.2.3 Text Data Entry

For the entry of text data from TDQ2 - difficulty accounts for 'difficult' and 'not so difficult' patients and comments from the last page of the questionnaire - no clerical help was available. This proved to be both a burden and a benefit. On the one hand, transcribing some 65,000 words of text was extremely time-consuming, on the other hand, there is no substitute for knowing one's material well and text entry and correction afforded a familiarity with the nuances of the written accounts which mere proof-reading and rating could not have matched.

German language texts from hand-written accounts were transcribed directly into a wordprocessor (MS Word 6.0), those from typewritten accounts (approx. 5% of the total) were scanned into an OCR program (Xerox Textbridge, which includes a German dictionary,) and corrected on-line. English language texts from hand-written accounts were dictated into a voice recognition program (IBM VoiceType) and corrected on-line from audio-playback; those from typewritten accounts (approx. 11% of the total) were scanned into Textbridge and corrected with a spell-checker. English translations of the German language texts were dictated into VoiceType and hardcopies were checked for accuracy of transcription and translation.

As the written accounts were generally more expansive than those in TDQ1 (with an average length of 277 words, more than double that of TDQ1, covering a range of 17 words to 869 words for the English sample), the time-consuming preservation of some of the contextual features of the accounts seemed less crucial. Paragraphs in the original text were therefore preserved but indentations and unusual positionings of text were not. Otherwise, the following transcription conventions were observed: Words which had been omitted from the original accounts but could be inferred from the context were entered into the transcripts in square brackets. Words which could not be deciphered were entered as '{?}' with each missing word being represented as a single question mark. Spelling errors were corrected but neologisms (for instance 'tizzing') were preserved. Most capitalisations (especially text in upper case) were preserved as were idiosyncratic uses of hyphenation and punctuation. Some commas were added where this helped to structure text. Common abbreviations (such as i.e. or e.g.) were preserved, specialised abbreviations (e.g. PTSD, OCD) were either expanded throughout the transcript or expanded on their first occurrence with the abbreviation cited in brackets. Shorthand expressions were expanded, ampersands and plus signs were transcribed as 'and' except in instances where they had a different meaning (such as '++' in place of 'very') when they were preserved. Annotations, such as explanations of references in one account to the other account by the same therapist, were added in square brackets and italicised.

5.2.4 Ratings of Difficulty Types: Reliability Study

The ratings of therapists' written accounts in terms of the three difficulty types - transient, situational and paradigmatic - provide the dependent variables for this part of the study. While the general reliability and validity of the coding system have been investigated in relation to the accounts collected with TDQ1, dependable ratings of the salience of each type for the accounts collected with TDQ2 are essential for further analyses. These were obtained by averaging the ratings of a group of judges which could be demonstrated to have high inter-rater reliabilities.

5.2.4.1 Rater Selection and Training

The reliability study of the ratings of difficulty accounts in TDQ1 had demonstrated the utility of the rating instructions for use with a larger, heterogeneous group of untrained raters. As reliability of ratings (i.e. magnitude of correlations of mean salience scores with a hypothetical 'true score') was the prime concern in this last stage of the current study, the logical consequence was to select a smaller, homogenous group of raters and to adopt a rater training procedure for the evaluation of the written accounts collected with TDQ2. Furthermore, using 'volunteers' as raters for TDQ1 had led to long time lags between distribution of materials and receipt of completed ratings (up to six months in one instance), had induced divers motivations in raters, and had generated conflicts between misgivings about the task and a sense of personal obligation to the author, all of which might well have had a detrimental influence on the reliability of the ratings. In contrast, external judges in this part of the study (with the exception of the thesis supervisor) were paid for their troubles, trained jointly on the same day, and were required to complete their ratings within three weeks of the training date.

5.2.4.1.1 Selection of Raters

The personality characteristics which distinguish reliable raters in the field of psychotherapy are poorly researched (Moras & Hill, 1991) but are most likely to play an important part with measures requiring a high degree of inference (Kiesler, 1973), such as the category system used in this study. For these measures a greater degree of clinical experience is also desirable[26], although more experienced clinicians may be less easy to train because they can be more rigid in their beliefs and constructs (Lambert & Hill, 1994). Given that the author and his supervisor were 'fixed' members of the rating group (for practical reasons but also because of their high interreliability in the ratings for TDQ1), the most promising strategy seemed to be to select additional raters who were similar to each other. There was, however, one important deviation from this strategy: As the selection of subgroups of most reliable raters for the accounts in TDQ1 had led to the exclusion of the only female rater in all three categories, it was thought important to include at least one female rater in the rating group for TDQ2 in order to monitor the possibility of the rating system having a gender bias.

Experience with ratings in respect of TDQ1 had demonstrated that a subgroup of three to four raters could achieve satisfactory reliabilities. Bearing in mind the possible need to check reliabilities with the author and/or the thesis supervisor excluded, three additional raters were appointed. In view of financial constraints, this was the smallest number that might be expected to constitute a reliable subgroup. All three were in their mid-thirties, were in advanced training as psychodynamic psychotherapists and had personal experience of conducting quantitative research and an active interest in psychotherapy research. One of them (R.III), is a male Clinical

---

[26] as demonstrated for the ratings of difficulty type in the reliability study for TDQ1 (see 4.2.4 above)

Nurse Specialist in Psychotherapy with degrees in Engineering and Psychology; the other two are Clinical Psychologists, one female (R.IV) and one male (R.V), who work half-time in specialist psychotherapy settings. All three had limited experience of supervising (mainly of generic mental health work) and were keen enough and, being NHS employees, impecunious enough to give up half a weekend and some additional free time for interesting work attracting a relatively low financial reward.

5.2.4.1.2 Rater Training

The training session was conducted in a work setting on a Saturday. Raters were asked to meet at 10 a.m. for coffee and subsequently were given a thirty-minute introduction to the study and its background. Particular emphasis was given to the idea that difficulties occur naturally and inevitably in therapy and to the view that for the therapist successful coping with difficulties may lead to gains in professional development, much in the same way as successful coping with difficulties in therapy may lead to gains in functioning for the patient. Because of the raters' background in psychology, some familiarity with established sources of unreliability in human judgements, such as central tendency bias, halo effect or leniency bias, could be assumed. Reference to these were made by way of a reminder in the training session but without further elaboration.

Raters were then issued with modified rating instructions and definitions of the difficulty types. As at least one of the raters in the group evaluating accounts from TDQ1 had shown a tendency to make judgements on the basis of (his own interpretation of) the category labels rather than the definitions, difficulty types were relabelled 'A' for 'Transient', 'B' for 'Situational' and 'C' for 'Paradigmatic'. The previous labels were shown in quotation marks in the written definitions and instructions, and raters were verbally directed to rely on the elaborated definitions

when making judgements about salience. During the discussion of training materials three further rating directions were given verbally:

1) If a difficulty has come about in the way described by category B - 'situational difficulty' and if it is further plausible or evident from the account that the therapist has brought about the difficult situation him/herself for reasons which are described by category C - 'paradigmatic difficulty'; then both categories should attract a salience rating.

2) If the difficulty has arisen because the therapist is troubled by sexual feelings for the patient/client, this would most commonly attract a salience rating in category A - 'transient difficulty'. There may be circumstances, however, where the patient has significantly contributed to the difficulty (for instance by being provocative), in which case category B should be rated as well/instead, or where the therapist's reaction is idiosyncratic (for instance unusually strong), in which case category C should be rated as well/instead.

3) Where a 'single choice' is not clear because two or three categories have jointly attracted the highest salience rating, decision-making may be helped by the following questions:

a) *Is the difficulty primarily located outside the therapist?* If so, the most likely single choice would be category B.

If the difficulty is primarily located inside the therapist, the next questions would be:

b) *Has the difficulty primarily come about because of the therapist's lack of knowledge, skills or experience?* If so, the most likely single choice would be category A.

c) *Is the therapist' s response idiosyncratic?* If so, the most likely single choice is category C.

Directions 1) and 2) were the result of ambiguities which had arisen in connection with TDQ1, direction 3) represented an initial outline of a decision-making tree which might be incorporated in a future rating manual.

One further clarification was prompted by the psychodynamic orientation common to all three raters. It was stressed that the 'lack of experience' referred to in the definition of category A does **not** pertain to experience, which a therapist might acquire over time, in better handling of a pervasive countertransference response (otherwise most 'paradigmatic' salience ratings would have been reclassified as 'transient').

In addition to the ratings specified in the directions, raters were also asked to make a judgement about 'codability', by placing either a '+' or a '0' or a '-' in the bottom right hand corner of each of the pages, denoting 'easy to code', 'average codability' and 'difficult to code' respectively. This appraisal, which had previously been used in a more differentiated form in the ratings of the 'difficulties' and 'coping strategies' taxonomies, was designed to permit analyses from which accounts consensually judged as difficult to code could be excluded.

Training materials had been selected from the accounts collected with TDQ1 on the basis of mean ratings calculated from the most reliable subset of raters in respect of each category. In addition to the means, 'disagreement scores' were calculated by averaging the absolute values of the differences between the scores of each of the raters in the subset and the mean scores, thus providing an index of unanimity among the groups of raters. One 'pure' instance of each of the categories was used for initial

discussion by selecting accounts which had attracted a unanimous score of '3' ('almost certainly salient') in the particular category and which showed low disagreement scores in the other two categories. The accounts chosen in this way were nos. 39, 35 and 25 from batch1 of TDQ1 ratings (see appendix 12). Two further accounts chosen for discussion were no. 25 from batch 1, which had a particularly high disagreement score on one category, and no. 23 from batch2 (see appendix 12), which showed high disagreement scores on all three categories and could therefore be expected to generate a debate amongst the current raters. After practising the rating system jointly on these accounts and comparing raters' individual evaluations with the 'criterion score' (consisting of the mean salience ratings established in the TDQ1 reliability study), raters were presented with a collection of 18 further accounts (nos. X03, 5, 16, 17, 22, 26, 27, 44, from batch 1 and nos. 8, 10, 15, 26, 30, 31, 32, 36, 50, 55 from batch 2) which they evaluated on their own. These difficulty descriptions - which represented about 10% of the number of accounts to be rated from TDQ2 - were selected on the basis of having low disagreement scores on at least one of the categories, regardless of the mean salience score on that category. This ensured that raters had the opportunity of evaluating accounts at all levels of all categories against a clear, consensual criterion.

While raters had lunch, their scores were entered into an SPSS data template containing the criterion scores which had been prepared in advance. The results, shown in the tables below, were then fed back to the group and deviations from the criterion scores were discussed case by case.

Table 5:9 - Raters' Category Means and Correlations with Criterion Scores

	Crit. *Mean*	**R. III** *Mean*	**r**	**R. IV** *Mean*	**r**	**R. V** *Mean*	**r**
A - 'transient'	*1.41*	2.28	**.70**	*1.22*	**.53**	*1.34*	**.57**
B - 'situational'	***2.13***	1.28	**.69**	*2.17*	**.72**	*1.61*	**.54**
C - 'paradigmatic'	***1.48***	.28	**.16**	*1.28*	**.46**	*1.44*	**.63**

Figures in the first column of Table 5:9 show the mean salience ratings of the 'criterion' groups, i.e. the mean ratings of the best subsets of raters in the reliability study of TDQ1 averaged over all 18 accounts. Columns 2, 4 and 6 show the corresponding means for each of the current raters. Columns 3, 5 and 7 show the Pearson correlations of each current rater's individual salience scores with the average scores of the criterion groups.

It is evident from a comparison of the means that R. III has a much higher mean on category A ratings than the criterion group and correspondingly lower means for categories B and C. However, correlations with the criterion scores are substantial for categories A and B but very low for category C, indicating that R. III used a baseline for the first two categories which differs markedly from the criterion, but judged variation from that baseline in a substantially similar way. For category C both baseline and variations from it differ markedly from those of the criterion groups. In the subsequent discussion it became evident that R. III had not made use of the notion that an idiosyncratic response is indicative of salience for category C - 'paradigmatic' and had therefore consistently underrated it. Taking this rule into account, he verbally arrived at scores which were much more similar to the criterion ratings.

R. IV had mean salience ratings which in all categories were very similar to the criterion and the correlations of her scores are very acceptable if somewhat variable.

R. V had a lowered baseline on category B, but the correlations of his scores with the criterion were good in all three categories.

	α Three Raters	α Best Two Raters	α incl. Criterion
A 'transient'	**.78**	**.78**	**.83**
B 'situational'	**.72**	**.72**	**.81**
C 'paradigmatic'	**.39**	**.57**	**.65**

Table 5:10 - Reliability Study TDQ2: Interrater Reliabilities During Training

Internal consistencies within the group of raters, expressed as Cronbach's alphas, are shown in Table 5:10. Figures in the first column show substantial three-rater alphas for categories A and B, but the alpha for category C is depressed because of the poor performance of R. III. The second column shows reliabilities for the best subset of two raters. For categories A and B these are nearly identical to the three-rater alpha - all three raters contribute to the reliability, but the effect of adding a third rater is lost in the rounding off of figures to two decimal points. Excluding R.III improves the reliability for category C. Although it is below those for the other two categories, an estimate of the five-rater alpha (including R.I and R.II), using the Spearman-Brown prophesy formula (Nunnally, 1978) would yield an acceptable reliability of .77. The third column gives an indication of the alphas that would have been obtained, had the current raters been part of the original rating group (by including the criterion score in the analysis as if it had been a fourth rater). The figures are likely to be an underestimate, since the criterion score was averaged from several raters (ranging from 2 for category C to 7 for category B). However, this seems preferable to using

all the original individual scores which entered into the criterion, since this would have led to effectively comparing groups with widely diverging numbers of raters (ranging from 5 to 10). Categories A and B show high (four-rater) reliabilities, whereas category C is approaching an acceptable alpha even though R.III is now included again (excluding him from this group would have yielded a three-rater alpha of .75).

Rater training then had resulted in two raters reaching satisfactory standards in all three categories, whereas one rater was performing well in respect of categories A and B (albeit using idiosyncratic baselines) but reaching a poor standard for category C. Moras & Hill (1991) suggest that, because of the possibility that perceptual biases may be relatively impervious to further training, only raters who have reached a satisfactory reliability overall should be retained; however, because of the professional and personal associations of all three raters with the author, such a practice would have been difficult. All raters were therefore retained for the main task. Because of practical constraints, R.III was not asked to complete a further set of training ratings after the discussion of his previous scores, thereby incurring the risk of his category C ratings having to be excluded from the ratings of TDQ2 accounts.

5.2.4.1.3 Effects of Rater Training on Reliabilities

In view of the time and effort expended on rater training, the question arises whether it had any appreciable impact. As no baseline was taken from raters before the training event, the impact of the thirty-minute introduction cannot be evaluated. It is, however, possible to compare the three-rater alphas derived from the training session with alphas from the same three raters derived from the reliability study reported below. Such comparisons are summarised in Table 5:11 below:

	Three Rater α Training Accounts	Three Rater α All TDQ2 Accts.	Three Rater α Selected Accounts
A 'transient'	**.78**	**.59**	**.64**
B 'situational'	**.72**	**.72**	**.79**
C 'paradigmatic'	**.39**	**.77**	**.87**

Table 5:11 - Comparison of Interrater Reliabilities During and After Training

The first column repeats the alphas obtained after analysing ratings of the training accounts which were already reported in Table 5:10. The second column shows alphas for the same three raters averaged over all the accounts contained in TDQ2 taken from the reliability study reported in 5.2.4 below. It appears that the discussion of the training ratings did correct R III's underrating of the 'paradigmatic' difficulty type which had previously suppressed the alpha for that category but did not improve reliabilities for the 'situational' difficulty type. The reliability for the 'transient' difficulty type has actually declined, though this is likely to be the result of the contextual information on length of therapists' practical experience (which had been available for the training accounts) not being included in the rating materials for TDQ2.

The figures in column 2 may, however, be an underestimate of training effects as they were derived from all accounts in TDQ2 whereas training accounts had been specifically selected from those which showed high agreement among the criterion group on at least one difficulty type. Reliabilities in column 3 therefore pertain to accounts which had been consensually judged by raters as having above average

codability (scores above 0; see 5.2.4.2 below) and which could therefore be expected to form a better group for comparison. Indeed, there are moderate improvements in reliabilities although the 'transient' category is still some way below the level of the training accounts.

In summary, it appears that rater training was successful in correcting gross rater error, was moderately successful in raising overall levels of reliabilities, but was unsuccessful in counteracting the effects of the exclusion of an important item of contextual information.

5.2.4.2 Rating Materials and Instructions

Following the procedure adopted for the reliability study of ratings of materials collected with TDQ1, difficulty accounts from the respondents to TDQ2[27] were again arranged into two groups: Batch I contained an array starting with the 'difficulty with a 'not so difficult' patient' of respondent no.1, followed by the 'difficulty with a 'difficult' patient' of respondent no. 2, and the 'difficulty with a 'not so difficult' patient' of respondent no.3 and so forth, whereas batch II had the obverse arrangement, starting with the 'difficulty with a 'difficult' patient' of respondent no.1, followed by the 'difficulty with a 'not so difficult' patient' of respondent no. 2, and the 'difficulty with a 'difficult' patient' of respondent no.3, etc. As in the reliability study of TDQ1, this produced two arrays in which each account was drawn from a different therapist, thereby avoiding dependency between cases, which might confound some statistical analyses, and providing a design in which findings could be 'replicated' between the two halves (batches) of the total data set. A total of 102 usable English-language questionnaires had been received (see 5.3.1 below). Three of

---

[27] Only English-language accounts were used in the study. German-language accounts have so far not been rated.

these arrived too late for inclusion in the rating materials, two further respondents did not provide written difficulty descriptions and a further three only reported one difficulty. Reliabilities were therefore calculated on Ns of 96 in batch I and 95 in batch II respectively.

As previously, the two batches were collated into rating booklets (see appendices 10 and 11) which presented one account at a time together with a template for the recording of ratings. Each difficulty description was supplemented by information about the sex, profession(s) and theoretical orientation(s) of the therapist rendering the account. Information about length of practical experience was omitted so as not to confound the subsequent investigation of its association with 'transient' salience scores which, if found to be positive, would provide evidence for the validity of the difficulty types. Two out of the five rating booklets started with batch II in order to balance for practice or fatigue effects.

The rating instructions relating to and supplied with the booklets (see appendix 5), were nearly identical to those used in the reliability study for the first batch of TDQ1 accounts described in 4.2.4.2 above. The main differences consisted of the relabelling of the difficulty types as 'A', 'B' and 'C' in order to avoid raters making judgements on the basis of the labels rather than the definitions, and incorporating the different method of eliciting accounts (difficulties with 'difficult' and 'not so difficult' patients instead of 'past' and 'present' difficulties). Other materials supplied consisted of a sheet with the headings and instructions under which accounts had been elicited in TDQ2 and a sheet with definitions for the difficulty types, now labelled 'A' (transient), 'B' (situational) and 'C' (paradigmatic).

5.2.4.3 Data Entry and Analysis

Raw scores for both batches of difficulty accounts were entered directly from the completed rating booklets into two templates in the SPSS data editor. Single-choice ratings were coded by adding the value of 5 to the salience score, e.g. coding '8' for a salience rating of '3' in a category which also attracted a single-choice rating. 'Codability' judgements were scored as '1', '0', and '-1' for the codes of '+', '0', and '-' respectively. Two pairs of working data files were derived from the raw score files, one containing codes of '1' for all single choice judgements and codes of '0' for all other ratings, the other containing salience ratings with the addition of '5' removed from the single-choice ratings. The reliabilities of salience ratings calculated as Cronbach's alpha (Cronbach, 1951) were analysed using the SPSS 'Scales' module, reliabilities of single choice ratings calculated as Cohen's kappa (Cohen, 1960) were analysed using the SPSS 'Crosstabs' module.

5.2.4.4 Interrater Reliabilities: Salience Ratings

As the analysis of interrater reliabilities eventually demonstrated that for every difficulty type in both batches all five raters contributed to the joint reliabilities, detailed information on the correlations between pairs of raters is omitted here. Key indicators are shown in the table below together with the five-rater alphas:

	TRANSIENT		SITUATIONAL		PARADIGMATIC	
	Batch I	Batch II	Batch I	Batch II	Batch I	Batch II
Mean	**1.5**	**1.5**	**1.7**	**1.6**	**1.5**	**1.6**
Range	.72	.72	.76	.61	.41	.57
Average Pairwise Correlation	***.32***	***.40***	***.48***	***.51***	***.57***	***.55***
Best Fit with Total Group	R I	R I	R I	R I	R V	R I
α	**.70**	**.76**	**.82**	**.83**	**.86**	**.86**

Table 5:12 - Reliability Study TDQ2: Interrater Reliabilities

The main columns show results by difficulty type, each subdivided for comparison into first and second batch of difficulty accounts. The first row gives mean values of salience ratings averaged over five raters. The homogeneity of means is striking, particularly when compared with the results from TDQ1 (see 4.2.4.4 above). Similarly, the differences between lowest and highest individual salience means - reported in row 2 - are much smaller than in the previous study. Rater training appears to have established similar baselines between raters within categories and between categories within raters. The drop in means for the 'situational' difficulty category between TDQ1 and TDQ2 is particularly noticeable. While it is theoretically possible that this is due to the changes in the enquiry format, there appears to be no convincing rationale why this category should have been differentially influenced in such a way and it seems more likely that the training procedure persuaded raters to pay equal attention to all three categories.

Values for average pairwise correlations calculated from all ten possible permutations of rater pairings are shown in the third row and are as expected higher than they were for the heterogeneous group of raters evaluating the accounts from TDQ1. The next row shows which rater's salience scores had the highest correlations with the results for the whole group, in other words who contributed most to the overall reliabilities. This was in all but one case R I, demonstrating that while raters have been shaped by rating instructions and training procedures to think about difficulty accounts in much the same way as the author, their ratings are in general somewhat less reliable (as one might have expected).

The last row shows the values for five-rater alphas. These are substantial although the reliability for 'transient' difficulty salience ratings in batch I is somewhat depressed. As the sequence of the presentation of batches in the rating materials was balanced, this is unlikely to be due to lack of practice or fatigue effects and therefore most likely represents unsystematic error. (Estimated single-rater alphas, calculated using the Spearman-Brown formula (Nunnally, 1978), range from .31 to .55 though one would expect the author's single reliabilities to be higher, given the argument outlined in the previous paragraph.)

Unidimensionality of ratings was established by principal component analyses of all 475 (5 raters x 95 accounts) salience ratings for each category in batch I and the corresponding 480 (5 raters x 96 accounts) ratings in batch II. In every case only one factor with an eigenvalue above 1.0 could be extracted, providing evidence that raters did indeed use common understandings of the concepts they rated.

In view of the efforts made in this part of the study to increase reliability by eliciting longer accounts, refining rating instructions, and carefully selecting and training raters, the improvements in alphas compared with the best subsets of the rater group

for TDQ1, especially those relating to the first batch of that study, are less than impressive. For the category of 'transient' difficulties this may in part be due to the omission of information about respondents' length of practical experience, which some raters highlighted as making their task more difficult in comparison to the accounts used in the training materials. However, overall it appears that despite the level of inference required, the definitions of the difficulty types and the associated rating instructions are in themselves robust and can potentially be used successfully even with shorter accounts and in the absence of specific rater training.

Improvements in rater reliabilities may however be achievable by screening out accounts which are ambiguous, vague or uninformative, on the basis of the 'codability' scores assigned by each rater to every account. Agreement for these scores was modest with five-rater alphas of .52 and .53 for Batches I and II respectively. These may well in part reflect the absence of any detailed directions on how to arrive at a rating, but it is also likely that raters are idiosyncratic in their reactions to specific accounts as well as in their general application of this measure. Nevertheless, consensual agreements on the ease or difficulty of coding particular accounts are likely to reflect not only raters' confidence in their judgements but also qualities of the narratives. A simple codability index was therefore derived for each account by summing individual judgements of ease of coding. The codability index could therefore theoretically range from '-5' for the most difficult-to-code account to '+5' for the easiest. The actual distributions of scores are given in the table below:

Codability Score	BATCH I	BATCH II
-5	1	0
-4	2	1
-3	4	8
-2	10	7
-1	15	8
0	24	24
1	14	22
2	9	14
3	13	6
4	3	6
5	0	0

Table 5:13 - Distribution of Codability Scores

One would expect interrater reliabilities to decline with decreasing codability scores. In order to investigate whether this expectation is borne out empirically, five-rater alphas were calculated for subgroups. batch I can most conveniently be broken down into three groups by values for the codability index:

a) **- 5 to -1** (N=32), b) **0 to 1** (N=38) and c) **2 to 4** (N=25).

Batch II lends itself to being split into four groups:

a) **-4 to -1** (N=24), b) **0** (N=24), c) **1** (N=22) and d) **2 to 4** (N=26)

The tables below show subgroup alphas per difficulty type for Batches I and II:

BATCH I :

Codability Score:	α TRANSIENT	α SITUATIONAL	α PARADIGMATIC
-5 to -1	**.56**	**.71**	**.74**
0 and 1	**.55**	**.80**	**.76**
2 to 4	**.85**	**.88**	**.96**

Table 5:14 - TDQ2 Batch I: Interrater Reliabilities Differentiated by Codability

BATCH II :

Codability Score:	α TRANSIENT	α SITUATIONAL	α PARADIGMATIC
-4 to -1	**.72**	**.55**	**.68**
0	**.73**	**.65**	**.82**
1	**.67**	**.84**	**.86**
2 to 4	**.84**	**.92**	**.93**

Table 5:15 - TDQ2 Batch II: Interrater Reliabilities Differentiated by Codability

On the whole the results support the prediction although there are differences between difficulty types. In summary one might say that high codability scores enhance reliabilities for the 'transient' category, that low codability scores suppress reliabilities for the 'paradigmatic' category and that there is a linear positive relationship between codability scores and reliabilities for the 'situational' category. It might be argued that some of these associations are the result of codability scores being confounded with one of the difficulty types; however, product-moment correlations between category salience means and codability scores show these to be nearly independent of each other, with the highest r values of .16 and .20 for batches I and II respectively occurring between codability and mean 'situational' scores. The

influence of codability scores on the reliabilities of salience ratings is therefore evident and needs to be borne in mind in the course of further data analyses.

In summary, the reliability study confirmed that the salience of the three difficulty types for the written difficulty accounts collected with TDQ2 can be evaluated with satisfactory reliability by a group of five raters. There is therefore reason to have confidence in the mean salience ratings, derived from averaging the five raters' judgements, as being serviceable dependent variables for further data analysis.

5.2.4.5 Associations between Difficulty Types

Having estimated the reliabilities for each of the difficulty types in both batches, mean salience ratings for each difficulty account are now available for use in further analyses. As for the ratings of the difficulties collected with TDQ1 (see 4.3.1 above), intercorrelations between difficulty types were calculated. These were subsequently corrected for attenuation to provide a better estimate of discriminant validity. The two tables below display intercorrelations for each batch of data. Reliabilities for each difficulty type (expressed as Cronbach's alphas) are shown in square brackets in the diagonal, uncorrected correlations in italics below the diagonal and corrected correlations in normal script above the diagonal.

BATCH I :

N = 95	TRANSIENT	SITUATIONAL	PARADIGMATIC
TRANSIENT	**[.70]**	**-.41**	**-.27**
SITUATIONAL	*-.31*	**[.82]**	**-.40**
PARADIGMATIC	*-.21*	*-.37*	**[.86]**

Table 5:16 - Intercorrelations between Difficulty Types for Batch I

BATCH II :

N = 96	TRANSIENT	SITUATIONAL	PARADIGMATIC
TRANSIENT	**[.76]**	**-.19**	**-.19**
SITUATIONAL	***-.15***	**[.83]**	**-.54**
PARADIGMATIC	***-.15***	***-.46***	**[.86]**

Table 5:17 - Intercorrelations between Difficulty Types for Batch II

All intercorrelations are negative, indicating good discrimination between the difficulty types. The negative associations with situational difficulty salience had been expected, as the results obtained with TDQ1 (see 4.3.1 above) had already indicated that raters will tend not to judge difficulties as being simultaneously located in and outside the therapist reporting the difficulty. The negative correlations between transient and situational difficulty saliences, however, are different from the previous results which showed the two types to be independent of each other. It is possible that rater training enabled judges to distinguish more clearly between them in the second study. Both sets of results confirm, that transient and paradigmatic difficulty salience are not confounded with each other, even though they both require difficulties to be located within the therapist reporting them.

5.2.4.6 Interrater Reliabilities: Single Choice Ratings

Alongside their salience ratings for difficulty types, raters had also been asked to chose for each difficulty account the one type that would best describe it (see 5.2.4.2 above). Whereas salience ratings treat difficulty types as continuous variables, single-choice ratings view them as binary variables - they are either selected or not selected for each account. This generates a different set of dependent variables whose

reliability can be estimated by calculating Cohen's kappas (Cohen, 1960) between pairs of raters.

Kappa values for pairs of raters for each of the difficulty types in both batches are shown in Tables Table 5:18, Table 5:19, and Table 5:20 below (figures below the diagonal in normal script relate to batch I, *figures above the diagonal in italic script relate to batch II*):

'TRANSIENT' DIFFICULTIES:

	R I	R II	R III	R IV	R V
R I	*	*.26*	*.28*	*.21*	*.34*
R II	.33	*	*.08*	*.26*	*.23*
R III	.39	.16	*	*.10*	*.20*
R IV	.37	.21	.07	*	*.24*
R V	.32	.29	.23	.32	*

Table 5:18 - Transient Difficulties: Pairwise Reliabilities for Single-Choice Ratings

Average Kappa (all raters): **.27 / *.22***

Average Kappa (R III excluded): **.31 / *.26***

'SITUATIONAL' DIFFICULTIES:

	R I	R II	R III	R IV	R V
R I	*	*.36*	*.48*	*.32*	*.27*
R II	.40	*	*.46*	*.36*	*.28*
R III	.44	.27	*	*.43*	*.32*
R IV	.41	.38	.41	*	*.24*
R V	.46	.41	.37	.45	*

Table 5:19 - Situational Difficulties: Pairwise Reliabilities for Single-Choice Ratings

Average Kappa (all raters): **.40 / *.35***
Average Kappa (R III / *R V* excluded): **.42 / *.40***

'PARADIGMATIC' DIFFICULTIES:

	R I	R II	R III	R IV	R V
R I	*	*.36*	*.40*	*.35*	*.38*
R II	.46	*	*.32*	*.46*	*.46*
R III	.42	.20	*	*.20*	*.28*
R IV	.42	.48	.21	*	*.43*
R V	.40	.45	.27	.51	*

Table 5:20 -Paradigmatic Difficulties: Pairwise Reliabilities for Single-Choice Ratings

Average Kappa (all raters): **.38 / *.36***
Average Kappa (R III excluded): **.41 / *.45***

R III seems to have operated different criteria from other raters in determining his single choices. Average kappa values can be improved in five of the six instances above by removing his ratings from the total. Only for single-choice ratings of 'situational' difficulties in batch II is the average kappa improved by excluding a different judge - R V. Given that kappa values are generally expected to be lower than reliability coefficients (Barker et al., 1994), the averages for the best subset comprising four raters do not compare badly with the average pairwise correlation for salience data referred to in Table 5:12 above.

This suggests that there is reason to have some confidence in the usability of an alternative set of dependent variables for further data analysis, namely categorical difficulty types based on four-rater subsets. Using the criterion of three out of four raters having to agree on a single choice, it is possible to classify 70 accounts (74%) in batch I and 61 accounts (64%) in batch II in this way. Using the more lenient criterion of a majority of raters having to agree on a single choice (thus allowing for 2-1-1 splits) leads to 79 accounts (83%) in batch I and 84 accounts (88%) in batch II being conclusively assigned to a single choice category. In the event, however, single-choice ratings were not used for the analyses reported in this thesis.

5.2.5 Numerical Data Entry

Data for the Familiarity / Depth of Emotional Impact / and Subjective Duration (FID) Scales were entered as raw scores from the questionnaires directly into two sets (German and English) of two files each (one for 'difficult' and one for 'not so difficult' patients) in the SYSTAT data editor. Blank responses were entered as missing data, ambiguous responses were assigned to the lower score, and two separate responses to the same item were averaged (unless the next item was blank, in which

case both items were entered as missing data). 'Checksums', which had been calculated by hand on the questionnaires by summing all scores into one grand total, were entered as a separate variable. After data entry, checksums were also generated through the SYSTAT data module and compared with the checksum variable. Eleven discrepancies between the four pairs of checksum arrays were followed up by comparison with the original questionnaire data, leading to the discovery of twelve data entry errors representing 0.18% of the 6578 datapoints checked in this way.

Items with reverse polarities were then recoded and subscale mean scores generated. The accuracy of the SYSTAT command files written for this operation was verified by running two dummy questionnaires, with extreme answers in both directions and checking that the programme generated maximum and minimum mean scores respectively.

Data entry for the Objective Duration Index is described in 5.3.3.2 below.

Data for the IIP-51 Scales (see 5.2.1.1.5.5 above) were also entered as raw scores into two sets (English and German) of three files each ('difficult' patients, 'not so difficult' patients, therapist self-report) in the SYSTAT data editor. As for the FID scales, blank responses were entered as missing data. A few respondents explained missing data with comments (such as "don't know"). In four instances these comments such as "yes and no", "sometimes yes, sometimes not" clearly indicated that both poles of the rating scale were applicable. In these cases blanks were coded at the midpoint of the scale. Ambiguous numerical responses were assigned to the lower score and two separate responses to the same item were averaged unless the next item was blank, in which case both items were entered as missing data. Again, checksums, which had been calculated by hand on the questionnaires by summing all

scores into one grand total, were entered as a separate variable. After data entry, checksums were generated through the SYSTAT data module and compared with the checksum variable. Fourteen discrepancies between the six pairs of checksum arrays were followed up by comparison with the original questionnaire data, leading to the discovery of five data entry errors representing 0.025% of the 19,737 datapoints checked in this way. Two items, which belong exclusively to the factor short form, need to have their polarities reversed before being included in the 'too open' subscale (Barkham et al., 1996). These were recoded prior to generating subscale scores for factors and circumplex octants. Subscale scores were computed by calculating averages of non-missing values (which effectively sets missing data to the average value of the scale to which the item belongs). In principle, this procedure could lead to the artificial inflation of subscale reliabilities; in practice, however, this risk is likely to be minimal as missing data accounted for only around 1% of responses and were quite evenly distributed among items. The accuracy of the SYSTAT command files written for these operations was verified by running a data set with 16 dummy cases, each of which maximised scores on one subscale while setting all others to zero, and checking the output for the expected extreme values. This led to the discovery and correction of a switch between two item labels in the command file.

SASB INTREX data (see 5.2.1.1.6 above) were entered and subsequently analysed using software (Benjamin, 1997) which had been made available via SPR-list, an internet mailing list for members of the Society for Psychotherapy Research. It replaces a previous, rather cumbersome, program (Benjamin, 1988b) which ran under DOS using fixed-format ASCII input, by a more user-friendly version running under Windows 95 and using Microsoft Excel spreadsheets as data entry format. Output for the short-form version is generated as three separate text files:

1) PICTURES: (mainly intended for clinical use) giving visual representations of all ratings together with best-fit curves and their interpretations as well as attack, control and conflict pattern coefficients and weighted affiliation and autonomy scores;

2) RESEARCH: providing a complete listing of raw scores, the above three pattern coefficients and weighted scores for each rating, in a format which can be read into a database or statistical package;

3) COEFF: listing for each rating all 21 pattern coefficients followed by the three composite attack, control and conflict pattern coefficients, in a format which can be read into a database or statistical package.

The software used for the analyses below has to be regarded as a beta-version, as three bugs needed to be detected and fixed before the program became operational:

a) raw score ratings of '100' take up three spaces in the input spreadsheet which is only designed for a column width of two spaces, leading to the output textfiles RESEARCH and COEFF being scrambled. This can be circumvented by coding raw scores of '100' as '99' and reconverting them when editing the output file;

b) the raw data file used by the program for subsequent processing is given the wrong name in the documentation and has to be renamed to be called up correctly;

c) the line-breaks created when opening the RESEARCH and COEFF text files in WORDPAD (as recommended) cause both SYSTAT and SPSS data import routines to misinterpret a tab-delimited ASCII file as fixed format, leading to problems in defining variables. This can be circumvented by importing the file

into a spreadsheet, adding a row for variable names, naming variables and then exporting the file into a statistical package.

After making the above corrections, the SASB software ran without problems. Raw scores were entered from the questionnaires into two copies (one for German and one for English data) of the template provided. The completed SASB data entry templates were saved as a separate spreadsheets and edited, so that only raw scores were left (excluding column headings, processing codes and text strings). These were checked for illegal entries by calculating check sums per data column and visually verifying that these were indeed multiples of tens as one would expect given that all legal SASB raw scores are in units of tens. In addition, the data for every tenth respondent were checked against the original questionnaires. No data entry errors were detected in the 8320 datapoints checked in these ways.

Data entry for the two CCRT grids was guided by two considerations: the necessity to use a format which would allow for the calculation of a large number of correlation coefficients and their easy transfer for subsequent analyses within the current study, and the desirability of having the data available in a format which would allow for their export into dedicated grid analysis software for future analyses outside the scope of the present study. As the best compromise between these two objectives, German- and English-language grid data were entered into two single (Microsoft Excel) spreadsheets with one 12 x 16 (elements x constructs) block per respondent. Relevant Pearson correlation coefficients were entered as formulae in arrays of single cells with their Fisher transformations computed in adjacent cells. Both sets of values were then consolidated across all respondents into a results spreadsheet for export into SYSTAT. In order to check the accuracy of data entry, marginal sums had been calculated by hand for all grids. Corresponding auto sum cells had been placed at the

end of all rows and columns of each data block in the spreadsheets, allowing for instant comparison between the two sets of sums during data entry. The very few errors which did occur were therefore trapped immediately and corrected on-line. Finally, total block sums were calculated and compared and as these tallied in all instances, there is reason to have confidence in the accuracy of the 18,240 datapoints entered and checked in this way.

5.3 Data Analysis and Results

Initial data analysis was carried out in sections, one measure at a time, with the resultant computed variables entered into composite work files for subsequent data analyses. The following sections keep to the same format but are preceded by a summary of measures which describe the sample and subsamples.

5.3.1 Sample Characteristics

The total number of questionnaires received at the end of the data collection period was 30 for the German-language and 104 for the English-language subsample. Two questionnaires from the English sample were excluded from the start, one because most responses were missing, the other because it did not meet the criteria of the study as both difficulty accounts related to couple therapy. (The total values for N, on which most of the following descriptors are based, are slightly lower because of missing data.)

As the sample cannot lay claim to representativeness through random selection of respondents, sample characteristics are especially important as indicators of generalisability. Summary statistics for respondents' characteristics are shown in the table below:

	English-Language Sample N = 102	German-Language Sample N = 30
Sex:	**F: 59.8% M: 40.2%**	**F: 56.7% M: 43.3%**
Age Mean (SD):	**45.2** (9.5)	**43.6** (8.0)
Median:	**46**	**43**
Range:	**26 - 71**	**32 - 60**
Practice Length Yrs.		
Mean (SD):	**12.8** (7.1)	**12.7** (8.3)
Median:	**12**	**10**
Range:	**1 - 33**	**1 - 34**
Practice Lgth. DP		
Mean (SD):	**10.6** (7.0)	**9.8** (8.8)
Median:	**9.5**	**7.0**
Range:	**.25 - 32.0**	**.25 - 33.9**
Practice Lgth. NSD		
Mean (SD):	**11.2** (7.4)	**9.3** (7.4)
Median:	**10.0**	**6.0**
Range:	**.25 - 32.3**	**.50 - 22.0**
Profession		
Psychologist:	**63%**	**57%**
Psychotherapist:	**45%**	**67%**
Counsellor:	**25%**	**---**
Nurse:	**12%**	**---**
Psychiatrist:	**11%**	**23%**
Social Worker:	**2%**	**---**
Arzt (Medical Doctor)	**---**	**20%**
Other Professions:	**9%**	**7%**
Theor. Orientation		
Analytic/Psychodyn.:	**3.8** *(96%)* (80%)	**4.0** *(100%)* (83%)
Humanistic:	**2.4** *(84%)* (55%)	**1.8** *(63%)* (40%)
Cognitive:	**2.3** *(84%)* (49%)	**1.6** *(70%)* (30%)
Behavioural:	**1.6** *(72%)* (30%)	**1.5** *(77%)* (33%)
Systemic:	**1.6** *(67%)* (28%)	**1.7** *(70%)* (33%)
Other Orientation:	**1.0** *(27%)* (19%)	**1.0** *(10%)* (7%)

Table 5:21 - Survey 2: Sample Characteristics

Looking first at the English-language sample (in column 1), it is apparent that the ratio of female to male respondents in the first row) is less evenly balanced than that for TDQ1 (see 3.3.1. above), but not alarmingly so.

Statistics for therapists' age - in the row below - show good variability and a considerable range. The same is true for years of practical experience at the time of filling in the questionnaire, shown in the next row down. Density and probability plots revealed values to be approximately normally distributed around a mean of nearly 13 years, with a slight right skew as indicated by the lower median. Both sets of figures attest that the typical participants in this study were mature and seasoned therapists who might be expected to express their accumulated life experience and clinical wisdom in the accounts of their difficulties. There is therefore no reason to be concerned that this study might rely too heavily on therapists in early career stages. Compared with TDQ1, one might say that typical respondents have matured in their clinical experience over the intervening two years at the same rate as the author and the study (a median of 12 years as compared to 10 years), while ageing somewhat more rapidly (a median of 46 as compared to 40 years).

Rows 4 and 5 of the table show descriptive statistics for years of practical experience at the time of the onset of difficulties with 'difficult' and 'not so difficult' patients respectively. While difficulties with not-so-difficult patients occurred on average marginally more recently than those with difficult patients, it is apparent that accounts of both difficulty types typically relate to the experiences of senior therapists.

The breakdown of professional background in row 6 shows that the sample is heavily weighted towards psychologists, with nearly two out of every three participating therapists reporting this designation. As respondents could endorse more than one category, percentages in this row of the table sum to more than 100. The designation 'psychotherapist' was most frequently endorsed in conjunction with another profession, for instance all nurses and more than a quarter of psychologists in the study also identified themselves as psychotherapists. A quarter of the sample call

themselves counsellors, but 60% of these also say that they are psychologists, indicating that there was a good response from counselling psychologists. About one in ten of the sample specified an 'other' profession; those specifically mentioned were: Lecturer (3 instances), Psychoanalyst (2 instances), Academic, Occupational Therapist, Probation Officer, Tutor (all cited once).

Details of respondents' theoretical orientations are shown in the final row. The figures in **bold script** show the mean influence of each orientation on a respondent's practice, rated on a six-point scale ranging from '0' ("not at all" influenced) to '5' ("very greatly" influenced). The next column of figures - in brackets and *italics* - shows the percentage of respondents who were influenced at all (ratings of at least '1') by each orientation, whereas the last column of figures - in brackets and normal script - shows the percentage of respondents who were strongly influenced (ratings of '3' and above) by each orientation. Percentages sum to more than 100 because of multiple endorsements. The predominant influence of psychodynamic theory on this sample is one of the striking features, another is the pervasive influence of all orientations (as shown by the percentages in the second column). As one might expect from a sample of experienced therapists (Skovholt & Rønnestad, 1995), theoretical purity has given way to an integration of theoretical influences - only 8 % of respondents claim to have only been influenced by a single school. More than a quarter also cite theoretical influences other than those suggested in the questionnaire, indicating the breadth of therapeutic convictions represented in the sample. Those specifically mentioned were: Existential (4 instances), Cognitive Analytic Therapy (3 instances), Personal Construct (3 instances), Integrative (2 instances), Neuropsychology (2 instances), Brief Therapy, Common Sense, Constructivist, Feminist, Group Analytic, Group Dynamics, Helliger/Beaumont, Interpersonal, Life Experience, Narrative, Non-

Directive, Powermapping, Self-Psychology, Transactional Analysis, Transpersonal, Trauma Incident Reduction (all cited once).

Descriptors for the much smaller German-language sample are shown in the second column of the table. Compared with the English sample, it is somewhat more evenly balanced for sex but shows a similar average age, albeit within a more restricted range. Mean practice length is practically identical, but there is a greater preponderance of less experienced therapists in the German sample, as indicated by the lower median. The statistics for practice length at the time when the difficulties occurred have to be treated with some caution, as they are based on even smaller Ns (23 and 19) because of missing data. With that proviso, it appears that German respondents reached further back into their memories when thinking of difficult situations, and tended to include events occurring at earlier career stages, than their English colleagues.

Within professional allegiances, the absence of nurses in the German sample can be accounted for by cultural differences (nursing is not a pathway to therapeutic practice in German speaking countries), whereas the absence of social workers and counsellors is most likely due to sampling restrictions. The category of 'Arzt(in)' in the German questionnaire reflects the situation in Germany and Austria, where a significant proportion of psychotherapeutic services are delivered by medical doctors with a 'Zusatztitel' (additional qualification) in psychotherapy (Strauss & Kaechele, 1998). Taken together with psychiatrists, and correcting for the overlap between the two categories, medical practitioners make up a third of the German sample; most of the other two thirds identify themselves as psychologists, and a few call themselves psychotherapists without any other designations. The two 'other' professions, each

mentioned once, were University Professor and University Assistant (temporary lecturer).

The breakdown of theoretical orientations shows an even stronger influence of psychodynamic ideas in the German-language sample, while cognitive and humanistic theories are somewhat less influential than they are among English respondents. The relative absence of 'other' orientations - only Group Therapy, Integrative and Interpersonal are each mentioned once - probably reflects the situation in Germany, Austria and Switzerland, where psychotherapy training is organised within theoretical schools rather than within professions, leading possibly to narrower allegiances with less scope for unusual and divers affiliations, compared to the UK.

In summary, the typical English respondent to TDQ2 was a therapist in their mid-forties, with 12+ years of practical therapeutic experience, mainly influenced by psychodynamic and humanistic theoretical orientations, and professionally identified as a psychologist and psychotherapist - much like the author of this study. In this respect, too, the definition of research as "in part, the discovery of the self through the detour of the other" (Hunt, 1989, p.42) seems to be borne out. The German-language sample appears broadly comparable to the English one. Some caution is indicated for analyses which might be influenced by the stronger medical representation in the German sample, but otherwise there is reason for having confidence in the validity of any cross-cultural comparisons.

Descriptive statistics for patient samples are reported in Table 5:22 below.

		English-Language Dataset		German-Language Dataset	
		DP (N=100)	**NSD** (N=97)	**DP** (N=29)	**NSD** (N=27)
Sex:		**F: 63%** **M: 37%**	**F: 73%** **M: 27%**	**F: 62%** **M: 38%**	**F: 71%** **M: 29%**
Age	Mean (SD):	**35.7** (9.4)	**35.8** (10.7)	**33.9** (9.3)	**36.9** (9.5)
	Median:	**35**	**35**	**30**	**35**
	Range:	**3 - 58**	**7 - 60**	**21 - 55**	**22 - 59**

Table 5:22 - Survey 2: Patient Sample Characteristics

The female/male ratio for difficult patients in both data sets is close to 2:1, that for not-so-difficult patients is closer to 3:1. The preponderance of females in the difficulty descriptions probably mirrors their stronger representation in therapists' caseloads. Male patients are somewhat more likely to be selected when therapists recall 'difficult' patients by both English- and German-speaking respondents. Further investigation showed that this tendency is shared by male and female therapists.

Age ranges in the English data set are influenced by the fact that three respondents reported difficulties from their work with children and adolescents. Patient mean ages are remarkably consistent across data sets and patient categories. Only difficult patients in the German data set show a slight right skew in the distribution of values.

Viewing both therapist and patient characteristics together, it appears that therapists typically reported difficulties which occurred 2 - 3 years ago with patients then about ten years younger than themselves.

5.3.2 Difficulty Types

The derivation of mean salience scores for the three difficulty types has already been described in 5.2.4 above. Before relating these to the main independent variables, their associations with other features of the data set, especially with the descriptor variables referred to in the previous section, are of interest. These will be examined in turn.

5.3.2.1 Difficulty Types and Patient Category

The most important of these relationships is with the category of 'difficult' vs. 'not so difficult' patient. The ways in which respondents described the patients with whom they were experiencing difficulties may well have been influenced as much by their general assumptions about the characteristics of 'difficult' or 'not so difficult' patients as by the specific qualities of the individuals they were calling to mind. Which of the two patient categories was involved in the situations they described is something that respondents were clearly aware of. However, they could not know which of the difficulty types were represented to what degree in their accounts. Associations between difficulty types and patient category could potentially confound associations between difficulty types and other variables. Conversely, finding associations between other variables and difficulty types in the absence of associations between the same variables and patient category, would present a more powerful argument for the importance of difficulty types.

Table 5:23 below shows the point-biserial correlations between the dichotomous patient-category variable and the continuous mean salience scores for difficulty types for each batch of TDQ2 accounts. Positive correlations indicate that salience is greater for 'difficult' patients, negative correlations that it is greater for 'not so difficult' patients.

	Paradigmatic	**Situational**	**Transient**
Patient Category Batch I	.01	.32	-.15
Patient Category Batch II	.01	.34	.14

Table 5:23 - TDQ 2: Correlations between Patient Category and Difficulty Types

In both data sets paradigmatic salience scores are effectively independent of patient category. In contrast, situational salience scores are significantly associated with patient category in each data set (n = 95; Bartlett chi-square statistic = 10.04; df = 1; p = .002 / n = 96; chi-square: 11.25; df = 1; p = .001), with respondents' difficulties more likely to be situational with difficult than with not-so-difficult patients. This might be expected given the definition of situational difficulties as located outside the therapist. The results for mean salience scores look more equivocal, showing a weak negative association with the 'difficult patient' category in batch I but a weak positive association in batch II. However, neither of these correlation coefficients is statistically significant, indicating that any association between the two variables is neither unequivocal nor reliable.

In summary one might therefore conclude that there is reason to be concerned that mean situational salience scores are confounded with patient category, whereas transient and paradigmatic salience scores appear to be independent of patient category. These results lend more weight to any relationships subsequently found between the salience of difficulties located within the therapist (transient and paradigmatic) and other variables.

As patient category was known to respondents, but difficulty types were opaque to them, a comparison between the two (treating them both as independent variables) might yield interesting information. The following associations of difficulty types with descriptor variables are therefore contrasted against patient category where appropriate.

5.3.2.2 Difficulty Types and Sex

The questions arising under this heading concern the main and interactive effects of therapist and patient sex on mean salience ratings of difficulty types.

Percentages of male vs. female therapists and patients have already been reported in 5.3.1 above. Table 5:24 below shows them in crosstabulation (absolute numbers are given in brackets). Because of the exclusion of missing data, marginal sums deviate slightly from those reported before.

PATIENTS:	THERAPISTS: Female	Male	
Female	**41.9 %** (80)	**25.7 %** (49)	67.6 % (129)
Male	**18.8 %** (36)	**13.6 %** (26)	32.4 % (62)
	60.7 % (116)	39.3 % (75)	*100 %* *(191)*

Table 5:24 - Survey 2: Therapist and Patient Sex Crosstabulation

As all therapists who contributed two difficulty accounts are represented twice in the table, their total N is almost double the number of respondents. Dividing the total sample into 'difficult patients' (n = 96) vs. 'not so difficult patients' (n = 95) yields almost identical percentages, indicating that there is no association between therapist/patient match with regard to sex and the likelihood of the patient being experienced as 'difficult' rather than 'not so difficult'.

Possible associations between therapist sex, patient sex, therapist/patient pairings and continuous mean salience scores for difficulty types were examined by calculating point-biserial correlations. The results are shown in Table 5:25 below. Correlation coefficients were computed per batch in order to preserve independence of observations.

	DIFFICULTY TYPE					
	PARADIGMATIC		SITUATIONAL		TRANSIENT	
	Batch I	*Batch II*	Batch I	*Batch II*	Batch I	*Batch II*
Therapist Sex	**-.01**	***-.10***	**.07**	***.03***	**-.29**	***-.12***
Patient Sex	**.08**	***.05***	**.08**	***.07***	**-.17**	***.18***
Therapist/Patient Pairings:						
Female / Female	**-.06**	***-.03***	**-.02**	***.17***	**.27**	***-.07***
Female / Male	**.09**	***.16***	**-.06**	***-.25***	**.02**	***.24***
Male / Female	**-.02**	***-.02***	**-.06**	***.01***	**-.12**	***-.11***
Male / Male	**.01**	***-.12***	**.18**	***.02***	**-.26**	***-.02***

Table 5:25 - Associations of Difficulty Type with Therapist and Patient Sex

Examining first the main effects of therapist and patient sex on difficulty types in the first two rows, it is apparent that the only consistent association across batches of any substance is that between therapist sex and transient mean salience scores. As males were coded '1', the negative sign indicates that transient type difficulties were more likely to be reported by female therapists. The association is quite modest (mean r = .21)[28] and only just reliable (p = .005 / p = .26 respectively; combined probability (see footnote 13) p = .05) but notable.

Reading across the table along the four possible permutations of therapist / patient pairings it is evident that there is very little consistency within difficulty types between batches and between difficulty types. Only for the female therapist / male patient pairing is there a hint that this match may be positively associated with difficulties located within the therapist, while being negatively related to difficulties experienced as stemming from external factors. This tendency is also evident, if both batches are amalgamated as shown below:

[28] Mean r was obtained after transformation of reliabilities into Fisher-z values and subsequent inverse transformation.

Therapist/Patient	DIFFICULTY TYPE		
	PARADIGMATIC	SITUATIONAL	TRANSIENT
Female / Male	**0.13**	**-0.15**	**0.13**
Probability	***0.08***	***0.04***	***0.07***

Table 5:26 - Associations of Difficulty Type with Female Therapist/Male Patient Pairings

Point-biserial correlation coefficients, shown in the first row, indicate weak positive associations between female therapist / male patient pairings and difficulties located within the therapist and a weak negative association with difficulties located in the patient or in external circumstances. The associated probabilities are shown in the second row. Although the assumption of independence of observations has been violated, this is mitigated by the fact that only in four cases were instances of female therapist / male patient pairings supplied by the same therapist across batches whereas the remaining 28 instances were supplied by different therapists.

In summary, the very modest association between therapist sex and transient difficulty type needs to be borne in mind for further analyses as a possible confounding factor. The suggestion that female therapists should show a tendency to report more difficulties with male patients which are located in themselves rather than in their patients or in external circumstances, is interesting and seems worthy of further investigation. It is, however, not well enough supported by the evidence presented here to merit consideration in further analyses.

5.3.2.3 Difficulty Types and Age

Although respondents' age and length of practical experience naturally are substantially positively associated ($r = .60$, $p < .001$; $r = .76$, $p < .001$ for the English- and German-language data sets respectively), a separate analysis of each variable is indicated as it might still reveal differential influences of professional and life experience. Both the effect of patient and therapist age and the effect of therapist/patient age differences on difficulty type are of interest. Therapist ages used in the following analyses are those at the time when the reported difficulty occurred, calculated by subtracting the answer to question 27 in the questionnaire ("How long ago was the onset of the difficulty?") from the stated therapist age at the time of responding.

The relationship of therapist age to patient category is already evident from Table 5:21. As there is little difference between years of practical experience at the onset of difficulties, there is also little difference between therapist ages at the onset of difficulties as the two variables are in a fixed relationship to each other[29]. The independence of therapist age from difficulty types, as expressed by product-moment correlations between age and mean salience scores near zero, is evident from Table 5:27 below:

	TRANSIENT		SITUATIONAL		PARADIGMATIC	
	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2
Therapist Age	**.07**	**-.06**	**.03**	**.01**	**-.04**	**-.07**

Table 5:27 - Associations between Therapist Age and Difficulty Types

[29] Point-biserial correlation coefficients between the dichotomous 'patient category' variable (difficult patient coded as '1', not-so-difficult patient coded as '0') and the continuous 'therapist age at onset of difficulty' variable are $r = .11$ for batch I, and $r = -.02$ for batch II.

Although there is no detectable influence of therapist age on patient category or difficulty type, the difference in ages of the therapist / patient pairings might exert such an influence. To investigate this possibility, age differentials were calculated by subtracting patient age from therapist age at the time when the difficulty occurred, resulting in negative values if patients were older than their therapists. Patients under the age of 18 were excluded, because of the assumption that for psychotherapy with children and adolescents the difference in ages would be a factor consciously taken into consideration by the therapist reporting the difficulty and by raters evaluating the account. Age differentials are approximately normally distributed around a mean of 6 years, with a range of -23 to 43 years.

The mean age differentials for difficult and for not-so-difficult patients are very close to each other; an independent samples t-test showed the difference between them not to be statistically significant ($t = -0.29$; $df = 92$; $p > 0.75$). The influence of age differential on difficulty types was investigated by calculating product-moment correlations between variables across the two batches of data. Results are shown in Table 5:28 below:

	TRANSIENT		SITUATIONAL		PARADIGMATIC	
	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2
Ther. / Patient Age Difference	**-.03**	**.10**	**.11**	**-.03**	**-.11**	**-.12**

Table 5:28- Associations between Therapist/Patient Age Differences and Difficulty Types

Although there is a hint that paradigmatic difficulties might be associated with smaller age differentials, the correlation coefficients are too small to be interpretable and are not statistically significant. Age differentials therefore give no reason for

concern as a variable which might confound results of further analyses involving difficulty types.

5.3.2.4 Difficulty Types and Practice Length

The absence of an association between patient category and therapists' years of practical experience at the time when the difficulty occurred has already been commented on above. The remaining question concerns the influence of practice length on difficulty type. In order to investigate this question, product-moment correlations between mean salience scores and length of practical experience at the onset of the difficulties were calculated for both batches of data. The results are shown in Table 5:29 below:

	TRANSIENT		SITUATIONAL		PARADIGMATIC	
	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2
Practice Length	**-.13**	**-.26**	**.04**	**.00**	**-.06**	**.03**

Table 5:29- Associations between Practice Length and Difficulty Types

Neither situational nor paradigmatic salience are related to practice length on the evidence of the respective correlation coefficients, which are all close to zero. Transient difficulty salience scores, however, show a modest negative association, indicating that there is a tendency for this difficulty type to occur earlier in therapists' careers. Calculation of combined probabilities shows this association to be quite reliable ($r = -.13$; $df = 90$; $t = 1.25$; $p = .11$ / $r = -.26$; $df = 87$; $t = 2.46$; $p = .008$ / $z = 2.59$; $p\ (comb) = .005$). One-tailed tests have been used because the relationship had been predicted in the specific hypothesis (see 5.1.2 above) that

- Transient mean salience scores will be negatively correlated with an index of the length of time a therapist had been practising when the difficulty first occurred.

This hypothesis has therefore been supported.

Two conclusions can be drawn from this association. One concerns the validity of the transience concept, which would lead one to expect difficulties of this type to be more prevalent in early career stages, as gaps in knowledge, skills and experience are among its defining characteristics. This link is borne out, even after removal of contextual information for raters to use in their judgements (see 4.2.2 and 5.2.4.2 above), which can be seen as an external validation of the transience concept. The second conclusion concerns the importance of professional as compared to life experience. Even though there is a substantial positive association between age and practice length, as indicated above, there is no association between age and transient difficulties, suggesting that the link between practice length and transience is entirely due to the professional component of development rather than to the acquisition of life experience. Being longer in the tooth does not in itself protect from transient difficulties; being seasoned as a practitioner does, albeit to a modest degree.

5.3.2.5 Theoretical Orientation, Professional Designation and Difficulty Types

Calculation of point-biserial correlation coefficients between the presence or absence of a professional designation and mean salience scores for difficulty types showed no consistent correlations above .15 in absolute value. The same was true for product-moment correlation coefficients between degree of influence of a theoretical orientation and mean salience scores for difficulty types. There is therefore no reason to be concerned that results of further analyses using difficulty types would be confounded by profession or orientation of respondents.

5.3.2.6 Summary of Results and Discussion

The preceding section investigated the relationships of the dependent variables derived from ratings of difficulty accounts (difficulty type salience scores) with the patient-category variable and other patient- and therapist-variables. The main aim had been to identify associations which might influence further analyses. The following results emerged:

a) 'Difficult' patient category shows a modest positive association with situational salience;

b) female therapist sex shows a weak positive association with transient salience;

c) length of practical experience at the onset of a difficulty shows a weak negative association with transient salience.

All three associations are reliable and need to be taken into account when interpreting results relating to the two difficulty types[30]. Results a) and c) are in line with expectations (result c) had been specifically predicted) and support the validity of the 'transient' and 'situational' difficulty definitions. The absence of any reliable associations with paradigmatic salience alleviates concerns that any future results relating to that difficulty type might be confounded with the variables investigated in this section.

---

[30] Therapist sex and practice length are not confounded. The point-biserial correlation coefficient between the dichotomous 'therapist sex' variable and the continuous 'length of practical experience at onset of difficulty' variable is r = .06.

5.3.3 FID Scales

Analysis of the FID Scale data served two main purposes:

a) to establish or confirm reliable subscales for each of the three variables 'familiarity', 'depth of emotional impact' and 'subjective duration', and

b) to derive subscale scores for each variable in respect of all difficulty accounts collected with TDQ2.

As outlined in 5.2.1.1.3.1 above, the 'Subjective Duration' scale was a new measure, specifically constructed for TDQ2, with thus far unknown psychometric properties for both English and German versions. Information about item characteristics and estimates of scale reliabilities for the English versions of the 'Familiarity' and 'Impact' Scales were available from TDQ1, though it was conceivable that the different mode of eliciting accounts in TDQ2 might have had repercussions for particular items (for instance by using past tense formulations for difficulties which are still ongoing). The German versions of these scales had only been piloted to establish that they were comprehensible to German speaking respondents. Their psychometric properties were unknown and one might have expected particular items to be affected enough by the translation to be at variance with the rest of the scale.

5.3.3.1 Total Scale and Subscales

As the three subscales hang together conceptually, giving rise to the expectation that extreme total scale scores would be related to extreme transient/paradigmatic salience scores for the associated difficulty accounts, the scale will in the following analysis initially be considered as a whole and the reliabilities and intercorrelations of the subscales will be investigated in a second step.

The allocation of items to scales is represented in the first three columns of Table 5:30 below; reverse polarity items are shown in italics.

			German Data						English Data					
			NSD (N=29)			DP (N=29)			NSD (N=98)			DP (N=97)		
FAM	IMP	DUR	$\bar{X}$	SD	*Rk*	$\bar{X}$	SD	*Rk*	$\bar{X}$	SD	*Rk*	$\bar{X}$	SD	*Rk*
	1		**3.9**	1.0	*23*	**4.5**	1.1	*22*	**3.4**	1.3	*17*	**4.0**	1.3	*20*
2			**3.6**	1.6	*19*	**3.0**	1.8	*12*	**3.7**	1.8	*18*	**3.0**	1.8	*10*
	3		**2.5**	1.7	*10*	**3.8**	1.7	*19*	**2.4**	1.9	*10*	**3.2**	1.8	*13*
		4	**2.8**	2.2	*13*	**2.2**	2.0	*7*	**2.5**	1.9	*12*	**2.4**	1.8	*7*
5			**1.7**	1.3	*3*	**1.7**	1.4	*1*	**1.3**	1.3	*1*	**1.4**	1.4	*1*
	6		**3.8**	1.3	*22*	**4.4**	1.3	*21*	**4.2**	1.2	*21*	**4.5**	1.3	*23*
		7	**2.3**	1.8	*8*	**3.7**	1.8	*17*	**2.2**	1.8	*7*	**3.3**	1.9	*14*
8			**2.6**	1.6	*12*	**2.2**	1.8	*6*	**1.8**	1.6	*4*	**1.6**	1.5	*3*
	9		**1.9**	1.8	*4*	**3.7**	1.7	*18*	**2.6**	1.9	*14*	**3.5**	1.7	*16*
10			**3.7**	1.7	*21*	**1.9**	1.2	*3*	**3.2**	1.9	*16*	**2.8**	1.8	*9*
		11	**2.2**	1.8	*6*	**3.6**	1.6	*15*	**2.1**	1.8	*6*	**3.4**	1.9	*15*
	12		**3.5**	1.7	*17*	**4.8**	1.4	*25*	**4.3**	1.6	*24*	**5.1**	1.4	*25*
13			**2.5**	1.6	*9*	**2.0**	1.6	*4*	**2.6**	1.8	*13*	**2.1**	1.7	*5*
	14		**3.0**	1.7	*14*	**4.3**	1.3	*20*	**3.1**	1.6	*15*	**4.1**	1.5	*21*
15			**3.7**	1.8	*20*	**2.7**	2.1	*9*	**4.2**	1.9	*23*	**3.6**	2.0	*17*
	16		**1.6**	1.6	*1*	**2.8**	1.5	*10*	**1.8**	1.5	*3*	**2.4**	1.7	*6*
		17	**3.6**	2.0	*18*	**4.7**	1.3	*24*	**4.0**	1.8	*20*	**4.9**	1.4	*24*
	18		**2.2**	1.8	*7*	**2.9**	1.6	*11*	**3.8**	1.3	*19*	**4.3**	1.3	*22*
19			**2.1**	1.6	*5*	**1.7**	1.5	*2*	**2.0**	1.6	*5*	**1.5**	1.5	*2*
		20	**1.6**	1.7	*2*	**2.3**	2.0	*8*	**1.7**	1.8	*2*	**2.4**	2.0	*8*
21			**4.8**	1.6	*26*	**4.7**	1.9	*23*	**4.3**	1.8	*25*	**4.0**	1.9	*19*
	22		**3.0**	1.4	*15*	**3.6**	1.2	*22*	**2.5**	1.6	*11*	**3.1**	1.6	*12*
23			**4.0**	2.1	*24*	**3.2**	2.6	*13*	**4.2**	2.1	*22*	**3.6**	2.1	*18*
	24		**2.6**	1.3	*11*	**3.3**	1.4	*14*	**2.3**	1.7	*8*	**3**	1.7	*11*
		25	**4.6**	1.7	*25*	**5.0**	1.5	*26*	**4.9**	1.3	*26*	**5.4**	1.2	*26*
26			**3.1**	1.6	*16*	**2.0**	1.9	*5*	**2.4**	1.7	*9*	**1.6**	1.6	*4*

Table 5:30 - TDQ2: Item Allocations and Descriptive Statistics for FID Scales

The statistics reported for each of the items are derived from four subsets of data based on accounts of difficulties experienced with difficult patients (DP) and not-so-difficult patients (NSD) supplied by German-speaking and English-speaking therapists. For each of these subsets **means** and standard deviations (SD) are shown

in the first two columns. Means relate to a seven-point scale (0 - 6) with 'zero' representing the pole thought to be indicative of 'transient' difficulties and 'six' representing the 'paradigmatic' end of the assumed spectrum. Reverse polarity items have been recoded to conform to this usage. The ***ranks*** (***Rk***) in the third column of each data subset relate to the means in ascending order (i.e. *RK 1* is assigned to the most 'transient' item).

Inspecting the data by columns to begin with, there appear to be adequately wide ranges of means, while extreme values - which would render an item ineffective - are absent. Inspection of normal probability plots confirmed that the distributions of means in all four data subsets approximate a normal distribution very closely. The magnitudes of standard deviations indicate that there is also good variability across items in all subsets.

Examining the data by rows suggests generally good consistency across the subsets both for means and for ranks. This is confirmed by the correlations summarised in the table below:

	German NSD	**German DP**	**English NSD**	**English DP**
German NSD	1	***.47***	***.82***	*.54*
German DP	**.55**	1	*.59*	***.85***
English NSD	**.89**	.58	1	***.80***
English DP	.63	**.85**	**.80**	1

Table 5:31- FID Scales: Product-Moment and Rank-Order Correlations between Samples

Figures in normal script below the diagonal show Pearson product-moment correlation coefficients calculated from the means of all 26 items per subset; figures in *italic script* above the diagonal show *Spearman rank order correlation coefficients*

calculated from the ranks associated with those means. Figures in **larger bold** typeface are those which allow for comparisons between language forms (within patient category) and within language forms (between patient categories), figures in smaller normal typeface are of little interest (commonalities and cross-comparisons) and are shown only for the sake of completeness. Both measures of association yield very similar results. Three out of the four associations of interest are high. Only the comparison of the two patient categories within the German data set shows more moderate, yet still substantial, correlation coefficients. Taken as a whole, these results provide evidence for the compatibility of results obtained from the scale across two different languages and two different contexts.

This overall picture may, however, obscure information about the consistencies of particular items which might provide useful information about translation weaknesses or specificity of items to patient category. An examination of rank differences across subsets shows some items to be highly consistent; for instance, item 25 is consistently the most highly endorsed duration item and has the highest mean in three out of the four subsets. Correspondingly, item 5 is consistently endorsed low and has the lowest mean of all familiarity items in all subsets. On the other extreme, item 18 is highly inconsistent in three out of the four comparisons. A partial explanation for this may be found in the formulation of the German item: "Diese Schwierigkeit wollte mir nicht aus dem Sinn" translating back into English approximately as "This difficulty was always on my mind" which is rather different from the actual English version "I felt concerned about this difficulty". This must in retrospect be regarded as a lapse in the questionnaire translation since a more literal version "Diese Schwierigkeit machte mir Sorgen" would make good sense in German. Furthermore, the German version is rather ambiguous about the context in which it can be located - outside sessions like

item 22 ("I was preoccupied by this difficulty outside the therapy sessions.") or during sessions - which might go some way towards explaining why this item is inconsistent within the German data set but quite consistent within the English one. It is of course possible that despite the inconsistencies item 18 may be a useful contributor to the 'Depth of Emotional Impact' scale, a question which is addressed below in the section on subscale reliabilities. No other items show stable discrepancies in ranks either within or between language groups, which may be taken as evidence that the German translation worked well for 25 out of the 26 items.

After the analysis of individual items, the reliability of the 'Familiarity', 'Impact' and 'Duration' subscales is the next logical focus of attention. A preliminary step for this analysis was to establish whether subscales contained more than one dimension by running Principal Component Analyses on all 12 data subsets in order to check whether more than one factor with an eigenvalue above 1.0 could be extracted. All subscales in both halves of the German data set were unidimensional, as were the Familiarity and Impact scales in both halves of the English data set. The English Subjective Duration subscale, however, split into two factors as shown in Table 5:32 below:

	'not so difficult' patient		'difficult' patient	
	Factor 1	**Factor 2**	**Factor 1**	**Factor 2**
item 7	**.88**		**.91**	
item 11	**.83**		**.78**	*.35*
item 20	**.82**		**.81**	
[α]	[.82]		[.82]	
item 17	*.49*	**.62**		**.92**
item 25		**.81**		**.90**
[α]		[.66]		[.82]
item 4		**.71**	*.33*	

Table 5:32 - Rotated Principal Component Loadings for English Subjective Duration Scale

Factor loadings were obtained after Varimax rotation. Loadings above **.5** are shown in **bold type**, loadings below *.5* are shown in *italics*, loadings below .3 are suppressed, and alphas for the resultant subscales are shown in square brackets.

Factor 1 forms a reliable subscale which is consistent over both patient categories. It contains one item each from the three subordinate constructs of subjective experience of absolute time which were used in the construction of the scale, i.e. 'Permanence', 'Flux' and 'Duration'. There is therefore good reason to retain this short subscale for further analysis of the English data sets as representing the construct of Subjective Duration as originally intended.

Item 4 was excluded from subscales because of its inconsistency, only showing a substantial loading for not-so-difficult patients. Factor 2 therefore forms a two item subscale, also remarkably reliable, particularly for the 'difficult patient' type. Items represent the 'Permanence' and 'Duration' constructs and are both of reverse polarity, raising the possibility that the direction in which items were answered might contribute to the two factor split. It is difficult to derive a convincing factor definition from these two items other than one like 'Brevity'. It seemed sensible not to discard this scale for the moment and to postpone decisions about its utility for further analyses until its associations with the other subscales had been examined.

The final step in this part of the analysis consisted of the investigation of the internal consistencies of subscales leading to decisions about the retention of individual items for the calculation of scale scores. To this end, Cronbach's alphas were calculated for each subscale for both patient types for both German and English data sets (six values for the German data set and eight values for the English data which contain two duration subscales).

Reliabilities for five out of the six unmodified subscales in the German data set are remarkably high, ranging from .80 to .91. The alpha for the 'Duration' subscale relating to accounts of 'difficult patients' is somewhat lower, but still substantial at .70. The alphas for the total unmodified scales are .85 and .78 respectively. From these results one may draw the conclusion that - despite the misgivings about item 18 raised above - the translation of the FID scale into German has been successful and has produced serviceable measures. These can be improved further, however, by the exclusion of a single item from each of the subscales - i.e. items 21 (Familiarity), 12 (Impact) and 4 (Duration). The reliabilities for the resultant final German scale versions are shown in Table 5:33 below:

GERMAN DATA SET:

	NSD (N=29)	**DP** (N=29)
FAMILIARITY (9 items)	**.93**	**.92**
IMPACT (9 items)	**.90**	**.90**
DURATION (5 items)	**.80**	**.75**
TOTAL SCALE *(23 items)*	***.87***	***.78***

Table 5:33 - Reliabilities of German FID Scales

The modified scales not only provide a way of calculating reliable scores for the analysis of data collected with TDQ2, but also form internally consistent instruments for measuring aspects of therapists' experiences of difficulties in the German language.

Reliabilities for the unmodified English Familiarity and Impact subscales range from .91 to .92. These could be only marginally improved by the exclusion of a single item for each scale - item 15 for the Familiarity subscale and item 12 for the Impact subscale. As the latter is congruent with the German data set, it was excluded from the final subscale version to provide consistency across languages. The Familiarity subscale was retained unmodified. The results demonstrate that the shortening of the scales from the longer versions contained in TDQ1 was successful. The reliabilities for the final scale versions are shown in Table 5:34 below.

ENGLISH DATA SET:

	NSD (N=97)	**DP** (N=98)
FAMILIARITY (10 items)	**.92**	**.91**
IMPACT (9 items)	**.91**	**.91**
DURATION (5 items)	**.73**	**.79**
DURATION 1 (3 items)	**.82**	**.82**
DURATION 2 (2 items)	**.66**	**.82**
TOTAL SCALE *(24 items)*	***.77***	***.78***

Table 5:34 - Reliabilities of English FID Scales

With the exception of Duration Factor 2 in respect of accounts relating to 'not so difficult' patients (which also appears to depress the reliability of the combined Duration scale for this patient category somewhat), reliabilities are high. English

subscale scores for use in further data analyses were calculated on the basis of these final versions.

5.3.3.2 The Objective Duration Index

The measure of the duration of a difficulty used in TDQ1, asking "how long did the difficulty last altogether?", had also been incorporated into TDQ2 (item 28), asking for the therapist's estimate expressed in months and weeks. As this question would not make sense for ongoing difficulties, the duration in weeks, if the relevant item (item 29) had been checked, was taken from the answer to the question "How long ago was the onset of the difficulty?" (item 27), which in most cases will represent an underestimate of the actual duration but is the best approximation available. Occasionally, answers to items 27 - 29 were inconsistent; for instance, one difficulty was described as 'ongoing' but the estimate of duration (derived from item 28) was shorter than the time elapsed since onset (derived from item 27). In such cases (totalling 10 out of 195 $\approx$ 5% in the English data set) the written difficulty account was used to decide which of the answers should be used.

Objective duration was encoded in weeks. As the answer format allowed respondents to estimate duration in a mixture of months and weeks, months had to be converted into weeks, using the convention that 1 month would be rounded down (to 4 weeks), and 2 months would be rounded up (to 9 weeks), while 3 months could be converted into 13 weeks without rounding. Multiples of these values were used to calculate higher numbers (for instance 8 months converts into 2 x 13 + 9 = 35 weeks).

The resulting 'Objective Duration Index' is of course objective only in as far as it uses actual time periods rather than therapists' subjective impressions of how long they endured a difficulty. As it is based on self-report data it could indeed still be

distorted by therapists' subjective perceptual biases. Summary statistics for the duration indices are shown in the table below:

	MEAN	SD	**MEDIAN**	***MIN***	***MAX***
GER NSD (N= 19)	**11.5**	23.4	**4**	*1*	*104*
GER DP (N=25)	**23.5**	32.9	**5**	*1*	*104*
ENG NSD (N=96)	**15.1**	32.0	**4**	*1*	*260*
ENG DP (N=97)	**30.4**	41.6	**17**	*1*	*260*

Table 5:35 - Objective Duration Index: Descriptive Statistics Across Samples

The samples in the rows are the German-language (GER) and English-language (ENG) data subsets each divided into difficulties with 'not so difficult patients' (NSD) and 'difficult patients' (DP). Two features are immediately apparent from this table:

The first is the right-skewness of all four distributions. Medians are much smaller than means, indicating the proportionally much higher number of low duration values. A log-transformation was therefore applied to the duration indices for further data analyses.

The second noticeable feature is the difference in mean difficulty durations between the 'difficult' and 'not so difficult' patient categories. This is statistically significant for the English data set (paired samples t-test, N= 93, t= -2.7, df =92, $p < 0.01$); because of missing data there are only 14 paired comparisons in the German data and the difference does not reach significance. How this association translates into correlations with the difficulty types is explored below.

The occurrence of the same maxima in both patient subsets of each language sample is coincidental; the values, which represent durations of two and five years respectively, are drawn from different respondents.

5.3.3.3 Scale Intercorrelations

Associations between the Familiarity and Impact scales and the ('objective') Duration index in TDQ1 have already been reported in 3.3.4 above. For TDQ2 the prompts eliciting the difficulty accounts have been changed (from 'past' and 'current or very recent' difficulties, to difficulties with 'difficult' and 'not so difficult' patients) and the ('Subjective') Duration scale has been added. Intercorrelations for the German and English data sets are shown in the following two tables. Figures in *italic*, above the diagonals, relate to not-so-difficult patients; figures in normal script below the diagonals, relate to difficult patients. The Duration scale in the English data set consists of the mean of all six items, analogous to the German data. The two shorter duration scales, derived from factor analyses (see Table 5:32 above), in the English data are labelled 'Dur. 1' and 'Dur. 2' respectively. 'Obj. Dur.' is the label for the log-transformed 'Objective' Duration index.

German Data Set:

	Familiarity	**Impact**	**Duration**	**Obj. Dur.**
Familiarity	1	***-.09***	***.10***	***-.02***
Impact	**-.37**	1	***.48***	***.14***
Duration	**-.19**	**.56**	1	***.74***
Obj. Dur.	**.08**	**.21**	**.47**	1

Table 5:36 - TDQ 2, German Data Set: FID Scale Intercorrelations

English Data Set:

	Familiarity	Impact	Duration	Obj. Dur.	Dur. 1	Dur. 2
Familiarity	1	***-.38***	***-.16***	***.01***	*-.22*	*-.08*
Impact	**-.32**	1	***.44***	***-.02***	*.53*	*.22*
Duration	**-.14**	**.46**	1	***.40***	*.85*	*.74*
Obj. Dur.	**-.05**	**.16**	**.42**	1	*.21*	*.35*
Dur. 1	**-.20**	**.51**	**.89**	**.27**	1	*.40*
Dur. 2	**-.14**	**.28**	**.51**	**.43**	**.18**	1

Table 5:37 - TDQ 2, English Data Set: FID Scale Intercorrelations

Viewing both tables together (the English partial table which corresponds to the German table is shown within heavier demarcations and in larger typeface), two sets of results stand out as notable. The first of these concerns the moderate negative association between the Familiarity and Impact scales, which runs counter to expectations as high scores on both of these scales had been thought to be indicative of paradigmatic difficulties. Though slight for German not-so-difficult patients, it is consistent and statistically significant ($p < .05$ / $p < .01$ / $p < .001$) across the other three subsets. These results clarify the associations obtained in TDQ1 with the same variables (see 3.3.4). It appears now that the absence of a correlation between 'past familiarity' and 'past impact' may well have been a result of the prompt "...a difficult situation...which occurred considerable time ago", presumably because it was not easy for respondents to recall accurately in retrospect how familiar a difficulty which happened a long time ago was at that time.

The second notable set of results concerns the intercorrelations with the new Duration scale. These are quite substantial both for the Impact scale and for the Objective

Duration index (which show insubstantial associations with each other). It appears that there is one aspect of the subjective impression of the duration of a difficulty which accords with the actual length of time for which it persisted, whereas another aspect corresponds to the emotional impact which the difficulty is having, linking the experience of its perseverance with the strength of its effect. For the English data set, the subscale Dur. 1 seems to tap the emotional impact dimension, while Dur. 2 is more closely related to actual length of time elapsed. While this distinction is obscured in the smaller German data set, it is, nevertheless, conceptually important enough to call the unified construct of 'subjective duration' into question.

The negative association between Familiarity and Impact scores is contrary to expectation and incompatible with the notion that the scales form a conceptually coherent entity which links them, together with duration, to the construct of paradigmatic difficulties. On the other hand, the alphas for the total FID scale across patient categories and language subsets range from .77 to .87, as detailed above, indicating that responses to each individual item are good predictors of responses to all other items. The most plausible explanation for this 'general factor' would be that it reflects individual response bias - the idiosyncratic use of the answering format - rather than a substantive factor common to all three scales. As discussed in 5.2.1.1.5.4 above, this general factor can be extricated by ipsatising scores - in this case by subtracting each participant's average response to all FID items from their individual item scores. Table 5:38 below shows the effect of this procedure on the scale intercorrelations of the English data set. Again, figures in ***italic***, above the diagonal, relate to not-so-difficult patients; figures in **normal** script, below the diagonal, relate to difficult patients. The Objective Duration index has been left out as its intercorrelations are unaffected by the removal of a constant from scale scores.

	Familiarity	Impact	Duration	Dur. 1	Dur. 2
Familiarity	1	***-.87***	***-.61***	*-.62*	*-.34*
Impact	**-.87**	1	***.16***	*.33*	*-.02*
Duration	**-.63**	**.18**	1	*.78*	*.67*
Dur. 1	**-.63**	**.29**	**.83**	1	***.20***
Dur. 2	**-.34**	**.13**	**.43**	**-.01**	1

Table 5:38 - TDQ2, English Data Set: Ipsatised FID Scale Intercorrelations

As a result of ipsatising, the Familiarity scale is shown to be strongly negatively associated with the other scales and the association between the Impact and Duration scales is weakened considerably[31]. Consequently, the theoretical construct of the FID scale, as previously conceptualised, now has to be abandoned. Instead, the possibility is raised that Familiarity might be inversely related to paradigmatic salience. Clarification of this issue has to be sought by investigating the relationships between the Familiarity, Impact and Duration measures and the difficulty types.

5.3.3.4 Associations with Difficulty Types

The relationships between experiential associates of a difficulty and its transient or paradigmatic nature were the concern of research question 2 (see 1.3 above) and led to the first general hypothesis outlined in section 5.1.2 above. This hypothesis can now - since the concept of a unified FID scale has been shown to be invalid - transformed into six specific hypotheses as follows[32]:

1. Therapists' mean Familiarity ratings for difficulties they report will be positively associated with mean paradigmatic salience ratings for those difficulties.

---

[31] The association between Impact and Duration. 2 has disappeared altogether, while the correlation between Impact and Dur. 1 survived, even though attenuated.

[32] Given the scale intercorrelations reported above, some of these hypotheses will, of course, be mutually exclusive - for instance as Familiarity and Impact are negatively correlated with each other, they cannot both be positively associated with paradigmatic salience.

2. Therapists' mean Familiarity ratings for difficulties they report will be negatively associated with mean transient salience ratings for those difficulties.
3. Therapists' mean Impact ratings for difficulties they report will be positively associated with mean paradigmatic salience ratings for those difficulties.
4. Therapists' mean Impact ratings for difficulties they report will be negatively associated with mean transient salience ratings for those difficulties.
5. Indices of the Duration of a difficulty will be positively associated with the mean paradigmatic salience rating for that difficulty.
6. Indices of the Duration of a difficulty will be negatively associated with the mean transient salience rating for that difficulty.

The actual relationships between English therapists' ratings of their difficulties in terms of the Familiarity, Impact and Duration measures, and independent raters' judgements of the accounts of these difficulties in terms of the salience of the three difficulty types (transient, situational, and paradigmatic), is summarised in Table 5:39 below:

	MEAN SALIENCE SCORES					
	TRANSIENT		**SITUATIONAL**		**PARADIGMATIC**	
	BATCH I	BATCH II	BATCH I	BATCH II	BATCH I	BATCH II
Familiarity	**-.25**	**-.17**	-.13	-.19	.01	.11
Impact	-.10	.11	.13	.10	**.25**	**.32**
Duration	.07	.15	.08	.22	.06	.10
Dur. 1	.04	.21	.10	.25	.10	.04
Dur. 2	.01	.03	.14	.09	.03	.07
Obj. Duration.	.13	.14	.17	.08	-.09	.11

Table 5:39 - TDQ2: Correlations between Difficulty Types and FID Scales

Adopting the criterion for 'noteworthy' associations outlined in 5.2 above (correlations above .15 in absolute terms and in the same direction across both batches), there are two such associations (shown in **bold** type). Both are consistent with the hypothesised relationships between FID scales and difficulty types.

One is the inverse relationship between transient salience and familiarity of a difficulty, which is also reliable ($r = -.25$; $df = 92$; $t = 2.47$ $p = .008$ / $r = -.17$; $df = 93$; $t = 1.64$; $p = .055$ / $z = 2.88$; $p(comb) = .002$ (one-tailed test)[33]). As predicted, unfamiliar difficulties are more likely to expose deficits in knowledge, skills or experience of the therapist who encounters them. This is particularly pertinent for therapists in earlier career stages - who are more likely to come across new situations, which they are unprepared for - since transience is also inversely related to length of practical experience (see 5.3.2.4 above). Hypothesis 2. is therefore supported.

The other noteworthy association, which is also reliable, is that between paradigmatic salience and the emotional impact of a difficulty ($r = .25$; $df = 92$; $t = 2.51$; $p = .007$ / $r = .32$; $df = 93$; $t = 3.26$; $p = .001$ / $z = 4.03$; $p(comb) < .0005$ (one-tailed test)). As predicted, idiosyncratic difficulties which are related to their internal conflicts or interpersonal problems are likely to register more strongly with the therapists experiencing them. Hypothesis 3. is therefore supported.

The remaining four hypotheses - nos. 1., 2., 5., and 6. - are not supported. There are also no reliable associations between FID measures and situational salience, for which no predictions had been made. The negative association between Familiarity and situational salience does, however, approach statistical significance ($r = -.13$; $df = 92$;

---

[33] As there are four comparisons involved (Familiarity, Impact, Subjective Duration, and Objective Duration), the criterion for statistical significance of the combined probabilities should be set as $p < .0125$ (.05 / 4).

t = 1.27; p = .204 / r = -.19; df = 93; t = 1.83; p = .07 / z = 2.17; p(comb) = .03 (two-tailed test)).

In summary, the evidence from TDQ2 does not support the hypothesised association between the paradigmatic nature of a difficulty and high levels of familiarity, impact and duration. Instead, it appears that low familiarity is characteristic of transient difficulties; that high emotional impact is characteristic of paradigmatic difficulties; and that duration, whether objectively or subjectively assessed, is not substantially or reliably related to any difficulty type.

5.3.3.5 Summary of Results and Discussion

The measures of Familiarity, Depth of Emotional Impact, and Duration were designed to capture specific qualities of the experience of a difficulty and to relate these to difficulty type. Analysis of data obtained with TDQ1 had shown the long forms of the Familiarity and Impact scales to be serviceable and highly reliable measures. Shortening of the scales (by more than a third for Familiarity and by more than half for Impact) for inclusion in TDQ2 did not adversely affect scale reliabilities, demonstrating the robustness of the measures. The newly designed Subjective Duration scale also had satisfactory reliability, despite being substantially shorter. All 26 items of the combined FID scale showed good variabilities, and their means within each scale were approximately normally distributed. Principal Component Analyses confirmed the unidimensionality of the Familiarity and Impact scales. Subjective Duration split into two components, both internally consistent, without overall scale reliability being substantially affected. In summary, there is reason to have confidence that the three scales are psychometrically sound.

Translations for the German versions of the scales - with the exception of one item - appear to have been quite adequate, as demonstrated by correlations between means and between ranks for all 26 items across patient categories and language subsamples. Reliabilities for the German Duration scale are satisfactory, those for the German Familiarity and Impact scales are high. Individual items in the combined scale perform as well as their English counterparts, and Principal Component Analyses confirmed the unidimensionality of all three scales. The German FID scales, too, appear to be psychometrically sound.

The Objective Duration Index (which has been shown to differentiate between patient categories) expresses therapists' estimates of the persistence of a difficulty in weeks. Respondents to TDQ2 in both language communities were able to use it and after log-transformation to correct for skewness it functioned as a serviceable complement to the FID scales.

From a technical point of view then, the study using TDQ2 has succeeded in establishing the FID scales and the Duration index as concise, reliable, sound, and cross-culturally validated measures of experiential concomitants of therapist difficulties.

From a conceptual viewpoint, the associations of the FID measures with each other and with difficulty types were of interest. Scale intercorrelations confirmed a negative association between Familiarity and Impact scales across all samples, while the Objective Duration index is in effect orthogonal to both scales. The Subjective Duration scale to some extent taps the same dimension as the Objective Duration index, as demonstrated by the correlations between the two measures (ranging from $r = .40$ to $r = .74$ depending on the sample). It is, however, also substantially associated with the Impact scale (correlation coefficients range from .40 to .56).

While ipsatisation of the scales demonstrates that the better part of this association is due to a general factor, the fact remains that the Subjective Duration scale (for the larger English sample) empirically splits into two components which are allied to the duration and impact elements. It seems therefore doubtful that the scale actually contributes to an enhanced understanding of the qualities of therapists' experience of difficulties. The ipsatised scale scores clarify, however, that the Impact and Subjective Duration scales are largely independent of each other, but are both negatively related to the Familiarity scale, while Objective Duration - though substantially associated with Subjective Duration - is near independent of the other two scales.

Conceptually then, familiar difficulties, regardless of their persistence, are affecting therapists less deeply. While it seems quite plausible that novel difficulties should have a higher impact than those which therapists are accustomed to, this finding invalidates the notion that paradigmatic difficulties should be both familiar and impactful.

Only two of the hypotheses derived from this now dysfunctional concept, concerning the relationships between FID measures and difficulty type, were empirically supported. Familiarity was shown to be a marker for transient difficulties, and Depth of Emotional Impact a marker for paradigmatic difficulties. The conceptual revision following from these findings parallels that of the transient - paradigmatic spectrum. While it had been originally thought that there was a single dimension of pervasiveness of a difficulty, which would be associated with levels of a coherent set of experiential qualities, it now appears that life isn't that simple. Instead, transient and paradigmatic salience are two quite distinct classifications of a difficulty, largely independent of each other (see Table 5:16 and Table 5:17 above), and each reliably

associated with one particular experiential quality. The third form of classification - situational salience - is not associated with either of these qualities. The third experiential quality - duration - is quite independent of the other two and of all three difficulty types, unless it is measured in a way which confounds it with impact.

What has emerged from the FID measures is, however, still conceptually coherent. Transient difficulties are more likely to be unfamiliar and to occur earlier in therapists' careers. This is consistent with their definition as arising from competency deficits which will be directly investigated in 5.3.7 below. Paradigmatic difficulties are likely to affect therapists more deeply. This is consistent with their definition as arising from therapists' enduring internal conflicts or persistent interactional patterns which will be investigated in 5.3.4, 5.3.4.4, and 5.3.7 below. The FID measures were not only technically successful, but have also contributed to the validation of the transient and paradigmatic difficulty type.

5.3.4 IIP Data

The data generated with the version of the IIP used in this study - the IIP-51 (see 5.2.1.1.5.5) - provide information about the range and strength of interpersonal problems reported by respondents in respect of themselves and their patients. They also allow for the third general hypothesis (see 5.1.2 above) to be operationalised and tested. Although the current study is primarily concerned with the associations between therapist and patient characteristics in relation to difficulty types, there is much to be learned from the levels of endorsement of total scales and their constituent subscales in the various versions of the IIP. The first steps in the analysis of the IIP data yielded by TDQ2 - before progressing to the associations with difficulty type - are therefore concerned with examining the prevalence of interpersonal problems

within the samples. It has to be borne in mind, however, that prevalence is recorded here entirely from the point of view of the therapists who completed the questionnaires. It is likely, that some of their evaluations of the interpersonal problems of their patients are generalisations from observations of in-session behaviours; indeed, some respondents have commented to that effect. While some clinical theories (such as psychodynamic or interpersonal schools) would postulate a close correspondence between interpersonal patterns in- and outside the consulting room, this cannot be taken for granted in the current context. Fortunately, this constraint on the interpretation of the IIP data is of little consequence for the issues under investigation, because it is precisely the therapists' points of view which are of interest here. Regardless of the accuracy of their evaluations of the interpersonal problems which distress their patients, it is the therapists' experiences which are of concern in this study. Their perceptions of the presence of these problems establish the foil against which the results of further steps in the data analysis, which are concerned with similarities and differences, can be contrasted and interpreted.

5.3.4.1 Prevalence

Data from the three specimens of the IIP-51 included in each questionnaire (therapist report on 'difficult' patient, therapist report on 'not so difficult' patient, and therapist self-report) were analysed according to the two conceptual frameworks, factors and circumplex, described above by computing eight factor-based and eight octant-based subscale scores, each calculated as the mean rating over the four items representing the subscale, for each inventory.

5.3.4.1.1 Subscales based on factor scores

For the factor based analysis, combined factors, as described by Barkham et al. (1994) were also calculated and labelled according to the psychosocial competencies proposed by Gilbert (1989). Where labels differ from those proposed by Barkham, the latter are quoted in brackets. For both sets of measures ipsatised versions were also calculated by subtracting individuals' mean item scores (see b) below) from the subscale scores.

In addition, the following indices were computed:

a) the grand mean for each inventory, calculated by summing all items, leading to

b) an item average, calculated by dividing total scores by the number of items (51), yielding a measure of individuals' general distress from interpersonal difficulties and allowing for comparisons between samples and comparisons with published norms from other populations;

c) an index of endorsed problems, calculated by a frequency count of items rated '1' or above, providing a measure of an individual's spectrum of interpersonal difficulties;

d) a positive item average, derived by dividing total scores by the number of endorsed items, giving an indication of the severity of experienced interpersonal problems.

Table 5:40 below shows the means for each of the factor based subscales and the computed indices across the three applications of the IIP, both in the untransformed and the ipsatised versions.

	SUBSCALE MEANS			IPSATISED SUBSCALE MEANS		
	DP (N=100)	NSD (N=99)	***T*** *(N=101)*	**DP** (N=100)	NSD (N=99)	***T*** *(N=101)*
Hard to be ASSERTIVE	**2.04**	2.38	***1.53***	**-.13**	.50	***.38***
Hard to be INVOLVED	**2.73**	1.90	***0.80***	**.56**	.02	***-.34***
Hard to be SOCIABLE	**2.62**	2.14	***1.16***	**.45**	.26	***.02***
Hard to be SUPPORTIVE	**2.54**	1.56	***0.90***	**.37**	-.32	***-.25***
Too AGGRESSIVE	**2.05**	1.46	***0.98***	**-.10**	-.42	***-.16***
Too CARING	**1.38**	1.81	***1.50***	**-.78**	-.07	***.36***
Too DEPENDENT	**2.12**	1.79	***0.97***	**-.03**	-.09	***-.17***
Too OPEN	**1.24**	1.17	***1.73***	**-.93**	-.71	***.59***
COMPETE	**2.05**	1.92	***1.26***	**-.03**	.92	***.54***
(Socializing) CO-OPERATE	**1.93**	1.65	***1.45***	**1.38**	.97	***-.57***
(Independence) ELICIT CARE	**2.43**	1.85	***.89***	**.59**	.11	***-.17***
(Nurturance) GIVE CARE	**1.96**	1.64	***1.20***	**1.15**	-.25	***-.61***
DISTRESS AVERAGE	**2.17**	1.88	***1.15***			
ITEMS ENDORSED	**41.1**	40.1	***34.9***			
POS. ITEM AVERAGE	**2.66**	2.38	***1.63***			

Table 5:40 - IIP-51, British Sample: Scale Indices and Factor Based Subscale Means

Figures in the first three columns show untransformed subscale means for 'difficult patients' (**DP**), 'not so difficult patients' (NSD) and therapists (***T***). Turning to the first eight rows, which represent the eight subscales based on four items each (but with

two items in the 'too open' subscale having had their polarities reversed - see 5.2.1.1.5 above), results show that, in general, respondents perceived their difficult patients as more interpersonally distressed than their not-so-difficult patients and themselves as least distressed. This is also clearly reflected in the distress averages[34] shown in row 13 and the positive item averages shown in row 15. T-tests show the means differences on these two averages to be statistically significant:

(Distress Average: DP vs. NSD: t= 4.31; df = 97; p < 0.001 / NSD vs. T: t = 11.66; df = 98; p < 0.001. Positive Item Average: DP vs. NSD: t = 4.43; df = 97; p < 0.001 / NSD vs. T: t = 13.26; df = 98; p < 0.001.)

This ordering is not surprising; after all "difficult people are people with difficulties", as the old counselling adage, which has helped many a beginning therapist survive their first problem client, would have it. There are, however, some interesting exceptions to this general trend. Problems with assertiveness are higher in not-so-difficult patients than in difficult patients (t = 2.53; df = 97; p = .012[35]); the same is true for problems involving being too caring (t = 3.23; df = 96; p = .002). Therapists, are more distressed by excessive openness than patients in either category (T vs. DP: t = 4.2; df = 99; p < .0005). These features disappear again, however, when subscales are combined into the four composite scales (see 5.2.1.1.5.3 above) shown in rows 9 - 12. These composite scales are shown `in a different typeface` because they were computed by taking the arithmetic means of the constituent scales, rather than by subtracting one from the other, which only makes sense for ipsatised scales (see below). Rather than providing information about the location of means on bipolar scales, the untransformed composite scales therefore provide information about

---

[34] It should be noted that distress averages differ somewhat from the arithmetic mean of the eight subscales because of recoding of the two reverse polarity items mentioned above.

[35] As there are eight comparisons among the subscales, this result is not in itself significant. It has been included because it is replicated in the German data set.

aggregate distress levels on secondary (higher order) constructs, regardless of polarity.

The ipsatised subscale means, represented by the figures in columns 3 - 6, show a more differentiated picture. Apart from rounding errors, values in the cells are congruent with the untransformed means from columns 1 - 3 after subtraction of the appropriate distress averages[36]. The composite scale means are of most interest here, as they are based on difference scores and therefore lend themselves to comparisons across columns. These were computed, following Barkham et al. (1994) by subtracting scores on the constituent scales with the format 'Too...' from scores on the constituent scales with the format 'Hard to...'. Resultant positive values therefore indicate predominance of problems at the deficit ('Hard to...') ends of the bipolar constructs, whereas negative values indicate greater distress arising from problems at the excess ('Too...') pole. Both difficult and not-so-difficult patients share marked problems on the co-operation or socialising dimension (H. Sociable - T. Open). What distinguishes them is that not-so-difficult patients are characterised by their problems with assertiveness - the deficit pole of the competition (H. Assertive - T. Aggressive) dimension, whereas difficult patients are characterised by their problems with supportiveness - the deficit pole of the nurturance or care giving (H. Supportive - T. Caring) dimension. Therapists describe themselves as suffering from a deficit in competitiveness (assertiveness) but from excesses in the other three, affiliative,

---

[36] It should be noted that the eight subscale means do not sum to zero, contrary to theoretical expectations. This can be explained by the cumulative effect of the difference in distress averages, caused by the recoding of the two reverse polarity items, being carried forward; and by the fact that individuals' mean item scores used in the transformations were calculated on the basis of all 51 items rather than on the basis of the 32 items forming each short version, as the former was held to be a more accurate estimate of individuals' general distress factors.

dimensions; the least problematic being independence or care eliciting (H. Involved - T. Dependent).

Viewing all these results together, it appears that therapists in this sample struggle with precisely those interpersonal features which might be seen as an asset in the therapeutic situation. They see themselves on average as too open, too caring and, to a lesser extent, as too dependent/overinvolved, while suffering from a lack of aggression and assertiveness. They describe all the patients who feature in their accounts of difficulties as having on average problems with co-operation, as not being open enough, which may be understood as indicating problems in forming and maintaining a working alliance. In addition, they describe their 'difficult' patients as being typically too unsupportive/uncaring, and therefore unlike themselves, whereas their 'not so difficult' patients typically have problems with aggression/assertion, just like themselves.

Finally, the number of items endorsed, shown in row 14, indicates that therapists report a significantly narrower range of interpersonal problems than they attribute to their patients in either category (NSD vs. T: $t = 4.93$; $df = 98$; $p < 0.001$).

Before considering the prevalence of interpersonal problems within the alternative framework of circumplex subscales, two comparison groups for the factor based results are of interest: published norms for clinical and non-clinical populations and the results from the smaller German sample.

	SUBSCALE MEANS				
	DP (N=100)	NSD (N=99)	T (N=101)	Non-clinical (N=143)	Clinical outp. (N=120)
Hard to be ASSERTIVE	2.04	2.38	1.53	1.12	1.87
Hard to be INVOLVED	2.73	1.90	0.80	0.91	1.37
Hard to be SOCIABLE	2.62	2.14	1.16	1.02	1.65
Hard to be SUPPORTIVE	2.54	1.56	0.90	0.65	0.96
Too AGGRESSIVE	2.05	1.46	0.98	0.84	1.49
Too CARING	1.38	1.81	1.50	1.25	1.72
Too DEPENDENT	2.12	1.79	0.97	0.90	1.60
Too OPEN	1.24	1.17	1.73	1.74	1.45
FULL SCALE	2.09	1.78	1.20	0.98	1.51

Table 5:41 - IIP-51: Comparisons of Factor Based Subscale Means with Other Samples

Table 5:41 above shows in the first three columns untransformed means from the present study, repeated from Table 5:40. Column 4 shows general population means and column 5 shows outpatient means, both taken from British studies reported by Barkham et al. (1996). Therapists' means are, as one would expect, generally quite similar to the values for the non-clinical group - somewhat higher overall with raised figures particularly for the H. Assertive, H. Supportive and T. Caring subscales, while showing a lower result only for the H. Involved scale. Means for both difficult and not-so-difficult patients are generally higher than those for the clinical comparison

group[37], the clearest exception being the T. Open subscale and, for difficult patients only, the T. Caring subscale.

The ipsatised mean scores on the bipolar scales, shown in Table 5:42 below, confirm and add to this picture. Therapists are close to the non-clinical sample on the three affiliative dimensions but well above them, and some way above the clinical sample, on the deficit pole of the competition dimension which is the one feature on which difficult patients differ from all other groups. Both difficult and not-so-difficult patients are well above the outpatient sample on the deficit pole of the co-operation dimension, lending weight to the notion that problems within the working alliance contribute to these patients being linked to difficulty situations in the minds of their therapists. Difficult patients are unique in showing a very high mean on the deficit pole of the nurturance construct, whereas not-so-difficult patients show problems in the opposite direction, but less markedly so than the remaining three groups. Finally, both difficult and not-so-difficult patients are located on the deficit pole of the care eliciting spectrum, while the outpatient sample shows the highest mean in the opposite direction, suggesting that dependent patients by and large do not give rise to difficulty situations, presumably because of their complementary fit to the excessive care-giving which their therapists are suffering from.

[37] It should be noted though that means for difficult and not-so-difficult patients are derived from therapist perceptions, whereas the comparison samples are derived from self-report.

	BIPOLAR SUBSCALE MEANS				
	DP (N=100)	NSD (N=99)	T (N=101)	Non-clinical (N=143)	Clinical outp. (N=120)
COMPETE	-.03	.92	.54	.29	.37
(Socializing) CO-OPERATE	1.38	.97	-.57	-.72	.18
(Independence) ELICIT CARE	.59	.11	-.17	-.09	-.23
(Nurturance) GIVE CARE	1.15	-.25	-.61	-.59	-.76

Table 5:42 - IIP-51: Comparisons of Bipolar Subscale Means with Other Samples

Table 5:43 below reports prevalence of interpersonal difficulties for the smaller German language sample in the same way as Table 5:40 above did for the English language sample, which has an N more than three times as large. On inspection of the two tables, the most striking feature is their close correspondence in most respects. Untransformed means for the eight four-item subscales are quite similar to the English sample, and overall distress and positive item averages (rows 13 and 15) are practically identical. The pattern of declining distress from difficult patients through not-so-difficult patients to therapists can again be observed in all combined scales (rows 9 - 12) and in the subscales, with the same exceptions as in the English sample. Therapists' perceptions of their own interpersonal problems and those of the patients with whom they experience difficulties do not appear to be influenced by cultural differences between Britain and German-speaking countries on the evidence of these measures. The ipsatised subscale means do, however, reveal one important distinction.

	SUBSCALE MEANS			IPSATISED SUBSCALE MEANS		
	DP (N=29)	NSD (N=29)	***T*** *(N=29)*	**DP** (N=29)	NSD (N=29)	***T*** *(N=29)*
Hard to be ASSERTIVE	**2.03**	2.58	***1.25***	**-.23**	.69	***.03***
Hard to be INVOLVED	**2.78**	1.98	***0.93***	**.52**	.09	***-.29***
Hard to be SOCIABLE	**2.85**	1.78	***1.03***	**.59**	-.10	***-.19***
Hard to be SUPPORTIVE	**2.58**	1.52	***1.06***	**.32**	-.37	***-.16***
Too AGGRESSIVE	**2.22**	1.47	***1.27***	**-.05**	-.42	***.05***
Too CARING	**1.61**	2.35	***1.66***	**-.65**	.46	***.44***
Too DEPENDENT	**2.21**	1.72	***1.17***	**-.05**	-.16	***-.05***
Too OPEN	**0.78**	1.29	***1.84***	**-1.48**	-.60	***.62***
COMPETE	**2.12**	2.03	***1.26***	**-.17**	1.11	***-.02***
(Socializing) CO-OPERATE	**1.82**	1.53	***1.44***	**2.07**	.50	***-.81***
(Independence) ELICIT CARE	**2.50**	1.85	***1.05***	**.57**	.25	***-.24***
(Nurturance) GIVE CARE	**2.10**	1.94	***1.36***	**.97**	-.83	***-.60***
DISTRESS AVERAGE	**2.26**	1.89	***1.22***			
ITEMS ENDORSED	**43.6**	41.2	***37.2***			
POS. ITEM AVERAGE	**2.65**	2.30	***1.63***			

Table 5:43- IIP-51, German Sample: Scale Indices and Factor Based Subscale Means

While the differences on the eight single subscales are similar if somewhat accentuated in comparison with the untransformed means, the bipolar scales organise these features into a clear pattern. Like their British colleagues, German-speaking

therapists struggle with excesses on the three affiliative dimensions and describe their difficult patients as having marked problems in care-giving, their not-so-difficult patients as suffering from an inability to assert themselves, and both patient categories as finding it hard to co-operate (although difficult patients show a much higher mean on this last dimension). In contrast to their British colleagues, however, German speaking therapists on average have no problems with assertiveness or aggression, scoring effectively near zero on the competition dimension. This finding makes intuitive sense, given that German culture, in my experience, generally places less value on mitigating devices in conversation, in contrast to English culture, where confrontation is more usually palliated by circumlocution, paraphrase and understatement (at least outside the boundaries of the Ridings). It would therefore not be surprising if therapists, who should be particularly sensitive to the conversational conventions of their culture, would amplify such a general pattern. More evidence for this conjecture might be found in German norms for the IIP which have, to my knowledge, as yet not been published (B. M. Strauss, personal communication, June 1998).

The finding does, however, shed more light on how therapists construe their not-so-difficult patients. While it was tempting to conclude from the English data set alone that therapists thought of not-so-difficult patients as being similar to themselves, the evidence from the German data set suggests that this is not the case and that the problems with assertiveness are in themselves one of the defining characteristics of this patient group. Not-so-difficult patients, on this understanding, are those who struggle to form and maintain a working alliance, but who do not challenge their therapists in an assertive manner and whose propensity for excessive caring may well enamour them to their therapists, offsetting some of their more troublesome

characteristics. They are therefore awkward enough for therapists to recall difficulties with them, but interpersonally inoffensive enough to be experienced as 'not so difficult'. Difficult patients, by contrast, are those whose problems in forming and maintaining a working alliance, are compounded by their also being experienced as unsupportive and self-centred.

5.3.4.1.2 Circumplex Subscales

Means for the eight subscales based on octants of the interpersonal circumplex (derived from a different 32-item subset of the IIP-51 than the factor-based scales) are shown in Table 5:44 below. If one pictures the interpersonal circle in the way in which it is usually represented, with the affiliation dimension on a horizontal axis where the left end denotes the 'cold', disaffiliative pole, and the dominance dimension on a vertical axis with the 'dominant' pole at the top end, then the octants in the table are ordered in clockwise succession, from 'Domineering' (N octant)[38] in row 1 to 'Vindictive' (NW octant) in row 8.

Turning to the means within groups and between subscales, it is apparent that, with one small exception, all follow a perfect circumplex ordering. There is a monotone succession from the highest down to the lowest value and then from the lowest up to the highest value again, following a pattern which approximates a sine wave[39]. The results therefore confirm that the shortening of the IIP by Soldz et al. (1995) was successful in preserving the circumplex ordering of subscales.

---

[38] The windrose metaphor for the position of octants is preferable to the clock metaphor because it gives names to the precise location of octant mid-points.

[39] Only in the therapist group is there a small deviation on the 'Exploitable' subscale, which shows a marginally lower mean than its neighbours, whereas theoretically one would expect it to be higher.

	SUBSCALE MEANS			IPSATISED SUBSCALE MEANS		
	DP (N=100)	NSD (N=99)	***T*** *(N=101)*	**DP** (N=100)	NSD (N=99)	***T*** *(N=101)*
DOMINEERING	**2.36**	1.49	***0.84***	**.19**	-.39	***-.31***
INTRUSIVE	**1.25**	1.11	***0.84***	**-.91**	-.77	***-.31***
OVERLY NURTURANT	**1.53**	2.09	***1.65***	**-.62**	.21	***.51***
EXPLOITABLE	**1.80**	2.21	***1.60***	**-.37**	.33	***.45***
NON-ASSERTIVE	**2.24**	2.49	***1.65***	**.07**	.61	***.50***
SOCIALLY AVOIDANT	**2.46**	2.05	***1.23***	**.29**	.17	***.09***
COLD	**2.81**	1.97	***0.86***	**.64**	.09	***-.29***
VINDICTIVE	**2.41**	1.56	***0.57***	**.24**	-.33	***-.58***
CIRCUMPLEX TOTAL	**2.11**	1.87	***1.16***			

Table 5:44 - IIP-51, British Sample: Scale Indices and Circumplex Subscale Means

What distinguishes the three groups is the location of the highest and lowest points. While untransformed and ipsatised means show the same picture (as they only differ by the value of a constant), the pattern is easiest to discern using the ipsatised scales in columns 4 - 6. Difficult and not-so-difficult patients share a common lowest point on the 'Intrusive' scale (NE octant), however, difficult patients show their highest group mean on the 'Cold' subscale (W octant), whereas not-so-difficult patients show their highest group mean on the 'Non-Assertive' subscale (S octant). Therapists' lowest group mean is on the 'Vindictive' subscale (NW octant), while their highest group means are diametrically opposite, on the 'Overly Nurturant', 'Exploitable' and 'Non-Assertive' scales (E, SE, and S octants). In summary, British psychotherapists locate the interpersonal problems of their difficult patients in the left, disaffiliative

half of the interpersonal circle and the interpersonal problems of their not-so-difficult patients in the bottom, subassertive half of the circle (with the latter representing a 90° anti-clockwise rotation of the former). They see their own interpersonal problems as occupying the lower right half of the interpersonal circle (a further 45° anti-clockwise rotation), mainly on a combination of the affiliative and sub-assertive poles of the underlying dimensions.

The patterns between groups and within subscales, looking across rows, confirm the above picture. Interpersonal distress means decline, from difficult patients through not-so-difficult patients to therapists, on five of the eight subscales. On the remaining three, which are the same scales which show the highest group means for therapists, the pattern is divergent. These results repeat the findings from the factor based subscales reported on above. What they lose in terms of differentiation on the affiliative dimensions, they gain in terms of coherence through the confirmation of the circumplex ordering.

Table 5:45 below shows the corresponding means from the German speaking sample. Again examining first the means within groups and between subscales, looking down columns, circumplex ordering is demonstrated with one minor exception ('cold' subscale, not-so-difficult patients). Turning next to the ipsatised scales in columns 4 - 6, the scores of the group of difficult patients are organised around the affiliative dimension of the interpersonal circle, with the highest value occurring on the disaffiliative pole ('Cold', W) and the lowest value occurring on the opposite pole ('Overly Nurturant', E). By contrast, the subscale scores of the group of not-so-difficult patients are mainly organised around the dominance dimension, showing the highest value on the subassertive pole ('Non-Assertive', S) and the lowest value on the upper right octant ('Intrusive', NE) closely followed by the octant centred on the

	SUBSCALE MEANS			IPSATISED SUBSCALE MEANS		
	DP (N=29)	NSD (N=29)	***T*** *(N=29)*	**DP** (N=29)	NSD (N=29)	***T*** *(N=29)*
DOMINEERING	**2.15**	1.13	***1.02***	**-.11**	-.76	***-.20***
INTRUSIVE	**1.07**	1.10	***1.12***	**-.19**	-.79	***-.10***
OVERLY NURTURANT	**1.72**	2.31	***1.67***	**-.55**	.42	***.45***
EXPLOITABLE	**1.76**	2.43	***1.48***	**-.50**	.54	***.26***
NON-ASSERTIVE	**2.37**	2.55	***1.41***	**.11**	.66	***.19***
SOCIALLY AVOIDANT	**2.78**	1.90	***1.10***	**.52**	.01	***-.12***
COLD	**2.80**	1.99	***1.04***	**.54**	.10	***-.18***
VINDICTIVE	**2.67**	1.43	***0.85***	**.40**	-.46	***-.37***
CIRCUMPLEX TOTAL	**2.17**	1.86	***1.21***			

Table 5:45- IIP-51, German Sample: Scale Indices and Circumplex Subscale Means

dominant pole ('Domineering', N). The lowest group mean for therapists occurs on the upper left octant ('Vindictive', NW) and their highest group mean is shown for the affiliative pole ('Overly Nurturant', E), diametrically opposite that of their difficult patients. In summary, German psychotherapists, like their British colleagues locate the interpersonal problems of their difficult patients in the left, disaffiliative half of the interpersonal circle (though excluding the 'dominant' octant) and the interpersonal problems of their not-so-difficult patients in the bottom, subassertive half of the circle. They see their own interpersonal problems as located in a segment defined by the nurturant and sub-assertive poles of the two dimensions. Minor discrepancies notwithstanding, there is a convergent picture emerging from both data

sets, in which therapists characterise their difficult patients as predominantly cold and hostile, their not-so-difficult patients as primarily lacking in assertion, and themselves as overly friendly and somewhat gullible. This meshes quite well with the results derived from the factor-based subscales. However, on the circumplex measure even German therapists struggle with problems of assertion, while British psychotherapists are somewhat prone to social avoidance[40].

5.3.4.2 Relationships between Subscales

The prevalence data reported above also allow for a direct comparison between the two types of subscales. This is of some general theoretical interest, given the debate over the adequacy of reducing interpersonal behaviour to two dimensions only (Gurtman, 1997), and of specific interest to this study as close correspondence between the two scale types would weaken the rationale for using a combined version of the IIP. Scale intercorrelations are reported in Table 5:46 below. Since the IIP was constructed as a self-report measure, only therapists' ratings of their own difficulties have been used. Correlation coefficients between scales sharing items are shown in **bold** script.

	Domin	**Intrusv**	**O-Nurt**	**Exploit**	**N-Assr**	**Avoid**	**Cold**	**Vindict**
H. Ass	-.27	-.22	-.08	.43	**.89**	-.26	-.19	-.38
H. Inv.	-.07	-.20	-.22	-.09	-.01	.10	**.65**	-.11
H. Soc	-.17	-.31	-.25	-.28	-.22	**.87**	.27	-.15
H. Sup	.29	.12	-.40	-.26	-.09	-.13	.25	**.50**
T. Agg	**.55**	-.04	.04	-.19	-.25	-.02	-.18	.23
T. Car	-.22	-.26	**.89**	.45	-.18	-.35	-.21	-.19
T. Dep	.13	-.06	.26	.14	-.05	-.07	-.39	.05
T. Opn	-.03	**.74**	.20	.20	-.18	-.16	-.14	.16

Table 5:46 - IIP-51: Correlations between Factor-Based and Circumplex Subscales

[40] This latter feature is not entirely counter-intuitive. I have heard myself quip on occasions that I became a psychotherapist not despite my social inadequacies but because of them, and have noted the chuckle of recognition amongst some of my colleagues.

In order to interpret the information in this table correctly, the shared variance accounted for by the common items between categories has to be taken into consideration.

Factor-based Subscale	Circumplex Subscale	Overlap Percentage	Shared Variance (r square)
T. Caring	Over-Nurturant	75%	79.2%
H. Assertive	Non-Assertive	50%	79.2%
H. Involved	Cold	50%	42.3%
T. Open	Intrusive	50%	54.8%
H. Social	Socially Avoidant	25%	75.7%
H. Supportive	Vindictive	25%	25.0%
T. Aggressive	Domineering	25%	30.3%
T. Dependent	Exploitable	0%	2%

Table 5:47 - IIP-51: Overlap and Shared Variance between Subscales

The results appear to confirm that factor-based and circumplex subscales tap different theoretical dimensions because in most cases nearly all the shared variance between the scales is accounted for by the items they have in common. In one instance (H. Involved vs. Cold) it even seems that the items specific to each scale are negatively related, thereby reducing the shared variance generated by the common items. The exception to this picture is provided by the H. Social and Socially Avoidant subscales which are very highly correlated (r = .87) even though they only share one item. The factor-based T. Dependent subscale shows only a weak association (r = .14) with the Exploitable octant, which should be its theoretical counterpart (see 5.2.1.1.5.5 above), having more in common (r = .45) with the neighbouring octant - overly nurturant. More interesting is its weak association (r = .13) with the Domineering subscale which speaks to the active, controlling aspect of dependency. Whether these patterns also hold true for a clinical population could in principle be easily established by analysis of the relevant items of full-scale (127 item) IIPs, collected from a clinical sample.

In summary, the decision to use an instrument which can be analysed both in terms of factor-based and of circumplex subscales appears to be vindicated by the results reported in this section

5.3.4.3 Relationships with Difficulty Types

Apart from establishing the general relationship between IIP scores and difficulty types, there is a specific hypothesis under investigation here, derived from the general hypothesis related to research question 3 (see 5.1.2 above). One would expect personal conflicts for therapists to be activated if they encounter patients whose patterns of interpersonal problems are like their own. The prediction is therefore that therapists are more similar to the patients with whom they experience paradigmatic difficulties than they are to those with whom they experience transient difficulties. This can be operationalised as follows:

- Indices of therapist / patient dissimilarity on the IIP-51 measure will be negatively associated with paradigmatic mean salience scores but not with transient mean salience scores.

As a first step, associations between difficulty types and IIP scores of the patients with whom the difficulties were experienced were investigated. Table 5:48 below shows product-moment correlation coefficients between paradigmatic (**PAR**), situational (***SIT***), and transient (TRANS) mean salience scores in the columns and factor-based IIP subscale scores, total scale scores (calculated from all 51 items), and positive item averages in the rows, across the two batches of the English data set.[41] As before, the criterion to determine whether an association was noteworthy was set at r $\geq |.15|$ consistently, and in the same direction, over both batches.

---

[41] Correlation coefficients for untransformed and ipsatised scores, which only differ by the value of a constant, are identical. A square root transformation was applied to the T. Caring and T. Open scores to reduce skewness of their distributions.

	DIFFICULTY MEAN SALIENCE SCORES					
	PAR		*SIT*		TRANS	
	B I	B II	*B I*	*B II*	B I	B II
DISTRESS TOTAL	**-.22**	**-.17**	***.14***	***.19***	-.01	.11
POS. ITEM AVERAGE	**-.29**	**-.06**	***.24***	***.20***	.00	.06
Hard to be ASSERTIVE	**.00**	**-.14**	***-.04***	***.02***	.05	.06
Hard to be INVOLVED	**-.18**	**-.04**	***.20***	***.17***	.02	.20
Hard to be SOCIAL	**-.12**	**-.08**	***.20***	***.04***	.04	.03
Hard to be SUPPORTIVE	**-.22**	**-.04**	***.21***	***.20***	-.03	.20
Too AGGRESSIVE	**-.21**	**-.08**	***.19***	***.16***	-.10	.14
Too CARING	**.02**	**-.03**	***-.18***	***-.05***	.04	-.13
Too DEPENDENT	**-.08**	**-.13**	***.03***	***.11***	.07	.13
Too OPEN	**.10**	**.07**	***-.14***	***.14***	.00	-.17

Table 5:48 - Correlations of Difficulty Types with Factor-Based Subscale Means

Starting with transience scores, in the last two columns, it is apparent that none of the associations meet the above criterion. For the first batch, the two total IIP scores are entirely unrelated to transience.

Turning next to paradigmatic salience, in the first two columns, the correlation coefficients with distress total meet the criterion. This negative association is reliable ($r = -.22$; $t = 2.18$; $df = 92$; $p = .03$ / $r = -.17$; $t = 1.68$; $df = 93$; $p = .10$ / $z = 2.7$; $p(comb) = .007$ (two-tailed test)). Among the subscales, there is also a clear trend towards negative associations, apart from the T. Caring and T. Open subscales. It appears that therapists describe patients they experience paradigmatic difficulties with as generally less interpersonally distressed. As therapists also rate themselves as lower on the IIP scales, there is a first indication here that they tend to view

themselves as rather more similar to the patients with whom they experience paradigmatic difficulties.

Looking finally at situational salience, in the middle two columns, there is a positive association, meeting the criterion, on one of the total scores, the positive item average, while the other one is very close to meeting it. The former correlation is also reliable ($r = .24$; $t = 2.36$; $df = 92$; $p = .02$ / $r = .20$; $t = 1.93$; $df = 93$; $p = .06$ / $z = 3.0$; $p(comb) = .003$ (two-tailed test)). Among the subscales, there are three (H. Involved, H. Supportive and T. Aggressive) noteworthy positive correlations. As situational difficulties are also conceptually and empirically related to difficult patients, the results are again consistent with the idea that difficult patients are patients with (interpersonal) difficulties. Furthermore, bearing in mind the lower scores of therapists on IIP measures, it appears that they see themselves as rather less similar to the patients with whom they experience situational difficulties.

Table 5:49 below shows correlation coefficients of difficulty type salience scores with circumplex subscales.

	DIFFICULTY MEAN SALIENCE SCORES					
	PAR		*SIT*		TRANS	
	B I	**B II**	***B I***	***B II***	B I	B II
CIRCUMPLEX TOTAL	**-.19**	**-.10**	***.11***	***.15***	.02	.10
DOMINEERING	**-.19**	**-.05**	***.23***	***.23***	-.11	.06
INTRUSIVE	**-.05**	**-.07**	***-.13***	***.13***	.11	.12
OVERLY NURTURANT	**.02**	**-.06**	***-.18***	***-.01***	.02	-.08
EXPLOITABLE	**.03**	**-.10**	***-.13***	***-.03***	.09	-.01
NON-ASSERTIVE	**-.08**	-.14	***.00***	***.05***	.10	.04
SOCIALLY AVOIDANT	**-.07**	**-.10**	***.17***	***.03***	.00	.03
COLD	**-.13**	**-.01**	***.18***	***.11***	-.03	.15
VINDICTIVE	**-.24**	**-.03**	***.32***	***.16***	-.14	.18

Table 5:49 - Correlations of Difficulty Types with Circumplex Subscale Means

The only noteworthy correlation coefficients, on the criterion adopted, are those between situational salience and the Domineering and Vindictive subscales, which define adjacent octants. However, the total scores[42] and some of the subscales seem to confirm the trend detectable in the factor-based measures: transience is unrelated to, paradigmatic salience tends towards negative association with, and situational salience tends towards positive association with IIP measures.

The next step in the analysis consisted of the construction of a suitable index of therapist / patient dissimilarity. The chosen measures here were the absolute differences between therapist and patient scores in each IIP composite index and subscale. An alternative measure might have been the squared differences - an index

---

[42] The negative association beween paradigmatic mean salience and the total distress score calculated from the circumplex items, although not meeting the criterion for being 'noteworthy', is nevertheless just reliable (r = -.19; t = 1.87; df = 92; p = .06 / r = -.10; t = .94; df = 93; p = .35 / z = 1.96; p(comb) = .05 (two-tailed test, no adjustment for multiple comparisons)).

related to the Euclidean metric (Gower, 1985). Krzanowski (1988) points out that both these measures are special instances of the Minkowsky metric, which exaggerates dissimilarity with increasing values of the integer to the power of which differences are raised. Absolute distances, therefore, appear to be the more conservative dissimilarity measure. However, Euclidean distances were tried as an alternative approach, but led in practice to skewed distributions unsuitable for calculating correlations. (Correcting these with square root transformations naturally produces simple absolute distances again.)

Table 5:50 below shows the product-moment correlation coefficients for each composite IIP index and each factor-based subscale, between mean salience scores for difficulty types and therapist/patient dissimilarities, both for untransformed and ipsatised scores, and for each of the two batches[43].

Looking for noteworthy associations - again on the criterion of $r \geq |.15|$ consistently, and in the same direction, across both batches - among the untransformed scores in the left half of the table, it is apparent that none of the correlations with transience (in columns 5 and 6) meet this standard. In fact, only half of them even show the same sign in both batches.

[43]In order to reduce skewness of their distributions, square root transformations were applied to H. Assertive, H. Supportive, T. Aggressive, T. Caring, T. Open, Compete, Co-operate, Give Care, Positive Item Average.

	UNTRANSFORMED SCORES						IPSATISED SCORES					
	PAR		*SIT*		TRANS		PAR		*SIT*		TRANS	
	B I	**BII**	***B I***	***BII***	B I	B II	**B I**	**BII**	***B I***	***BII***	B I	B II
DISTRESS TOTAL	**-.24**	**-.30**	***.02***	***.22***	.00	.14						
POS. ITEM AVERAGE	**-.27**	**-.31**	***.05***	***.27***	.12	.16						
Hard to be ASSERTIVE	**-.07**	**-.21**	***.04***	***.20***	-.11	-.05	**-.23**	**-.10**	***.27***	***.13***	-.10	-.13
Hard to be INVOLVED	**-.20**	**-.09**	***.19***	***.18***	.00	.07	**-.16**	**-.04**	***.22***	***.11***	-.02	.05
Hard to be SOCIAL	**-.27**	**-.16**	***.20***	***.30***	-.01	.08	**-.04**	**.06**	***.07***	***.16***	.07	.19
Hard to be SUPPORTIVE	**-.20**	**-.03**	***.15***	***.22***	.01	.10	**-.16**	**.14**	***.20***	***.10***	.07	.06
Too AGGRESSIVE	**-.20**	**-.20**	***.30***	***.20***	-.14	.08	**-.04**	**-.02**	***.02***	***.17***	-.00	-.01
Too CARING	**-.06**	**.24**	***-.07***	***-.13***	.08	-.02	**-.28**	**.08**	***.26***	***-.03***	.03	.05
Too DEPENDENT	**-.22**	**-.29**	***.09***	***.18***	-.08	.11	**-.12**	**.01**	***.07***	***.00***	.01	-.13
Too OPEN	**.08**	**-.07**	***.09***	***.08***	.10	.14	**-.16**	**-.09**	***.02***	***-.03***	.09	.11
COMPETE	**-.21**	**-.30**	***.19***	***.31***	-.13	-.00	**-.19**	**-.06**	***.21***	***.17***	-.08	-.11
CO-OPERATE	**-.26**	**-.15**	***.20***	***.27***	-.00	.13	**-.16**	**-.04**	***.06***	***.08***	.08	.23
ELICIT CARE	**-.27**	**-.19**	***.19***	***.22***	-.05	.09	**-.21**	**.02**	***.26***	***.07***	.01	-.05
GIVE CARE	**-.19**	**.11**	***.08***	***.07***	.06	.09	**-.26**	**.14**	***.28***	***.02***	.04	.06
FACTOR TOTAL	**-.29**	**-.17**	***.22***	***.31***	-.04	.10	**-.29**	**-.00**	***.31***	***.16***	.02	.07

Table 5:50 - Correlations between Difficulty Types and Factor-based Difference Indices

In contrast, associations with paradigmatic salience are consistently negative in all but two cases. The two full scale measures (based on all 51 items), in the first two rows, meet the criterion, as does the total measure of all factor-based subscales in the last row. Among the subscales, H. Social, T. Aggressive and T. Dependent meet the criterion, while H. Assertive, H. Involved and H. Supportive fail to meet it because of negative correlation coefficients smaller than -.15 in one of the batches. The T. Caring

and T. Open subscales show associations in opposite directions across the two batches. These last two results were already foreshadowed by therapists' high self-ratings and lower ratings of the patients with whom they experience paradigmatic difficulties, as noted above. The aggregate scales (mean scores on untransformed bi-polar scales) in rows 11- 14 show this picture from a different angle: the negative associations of paradigmatic salience with the Competition, Co-operation, and Care-eliciting scales all meet the criterion; the association with Care Giving is equivocal. Overall, the prediction that therapists would see themselves as more similar to patients with whom they experience paradigmatic difficulties than to those with whom they experience transient difficulties, appears to be borne out by these measures.

Turning next to the associations with situational mean salience scores, in columns 3 and 4, the most striking feature are the consistently positive correlations (indicating dissimilarity between therapists' perceptions of their patients and of themselves) on all measures, apart from the T. Caring subscale where the direction is reversed. Of the total scale measures, only the total based on factor subscales, shown in the last row, meets the criterion. Among the subscales, four (H. Involved, H. Social, H. Supportive, and T. Aggressive) meet the criterion; and among the aggregate scales the same three (Compete, Co-operate, and Elicit Care) which are noteworthy for their negative correlations with paradigmatic salience, also meet the criterion in respect of their positive correlations with situational salience. The finding that therapists view themselves as dissimilar from the patients with whom they are experiencing situational difficulties, at least on some dimensions, had not been expected. It does, however, tally with the finding noted above, that these patients have raised

interpersonal problem scores and are therefore dissimilar to therapists, who rate themselves generally low on these measures.

Looking at the correlations with ipsatised IIP scores, in the right half of the table, similar trends are still discernible, though much less consistent and rather attenuated. None of the correlations with paradigmatic salience and only two of the correlations with situational salience meet the criterion, and the consistency of the directions of associations across batches are no better for paradigmatic salience than they are for transient salience. It appears that a substantial proportion of the similarity between therapists and 'paradigmatic difficulty' patients is accounted for by the similarity in general distress scores. Removing this factor from the scores attenuates the similarities to the point of disappearance. There is, however, no reason to assume that therapists should show a consistent response bias in respect of themselves and their 'paradigmatic difficulty' patients but not in respect of their 'transient difficulty' patients. The 'general factor' therefore needs to be interpreted substantively, as a major component accounting for the similarities.

The next question arising concerns the reliability of the observed associations. While combined probabilities can be computed as before, the number of comparisons increases the expected incidence of Type I errors. Variables in Table 5:50 fall into three categories: General indices (3), composite subscales (4), and factor-based subscales (8). Allowing for the number of comparisons which each of the categories involves, the following criteria for statistical significance seem indicated: General indices, $p = .016$ (.05/3); composite subscales, $p = .0125$ (.05/4); and factor-based subscales, $p = .006$ (.05/8). Table 5:51 below shows consistent correlations with paradigmatic and situational mean salience scores for each of the untransformed IIP indices, together with associated probabilities and combined probabilities. In the

absence of specific predictions, two-tailed tests have been used for situational salience. Combined probabilities meeting the significance criteria are displayed **bold** with asterisks.

	PARADIGMATIC MEAN SALIENCE SCORES					**SITUATIONAL MEAN SALIENCE SCORES**				
	r		*p*		**p** comb. (one-tail)	**r**		*p*		**p** comb. (two-tail)
	B I	**B II**	*B I*	*B II*		**B I**	**B II**	*B I*	*B II*	
DISTRESS TOTAL	**-.24**	**-.30**	*.011*	*.002*	* **<.0005**	**.02**	**.22**	*.887*	*.036*	.11
POS. ITEM AVERAGE	**-.27**	**-.31**	*.005*	*.002*	* **<.0005**	**.05**	**.27**	*.633*	*.009*	.027
Hard to be ASSERTIVE	**-.07**	**-.21**	*.246*	*.025*	.030	**.04**	**.20**	*.728*	*.056*	.10
Hard to be INVOLVED	**-.20**	**-.09**	*.025*	*.188*	.022	**.19**	**.18**	*.065*	*.078*	.016
Hard to be SOCIAL	**-.27**	**-.16**	*.005*	*.065*	* **.002**	**.20**	**.30**	*.055*	*.003*	* **.001**
Hard to be SUPPORTIVE	**-.20**	**-.03**	*.028*	*.378*	.058	**.15**	**.22**	*.159*	*.037*	.016
Too AGGRESSIVE	**-.20**	**-.20**	*.031*	*.026*	* **.003**	**.30**	**.20**	*.004*	*.061*	* **.001**
Too CARING	**-.06**	**.24**				**-.07**	**-.13**	*.485*	*.230*	[.35]
Too DEPENDENT	**-.22**	**-.29**	*.018*	*.003*	* **<.0005**	**.09**	**.18**	*.411*	*.085*	.072
Too OPEN	**.08**	**-.07**				**.09**	**.08**	*.376*	*.457*	.25
COMPETE	**-.21**	**-.30**	*.022*	*.002*	* **<.0005**	**.19**	**.31**	*.069*	*.003*	* **.001**
CO-OPERATE	**-.26**	**-.15**	*.006*	*.073*	* **.002**	**.20**	**.27**	*.056*	*.009*	* **.001**
ELICIT CARE	**-.27**	**-.19**	*.005*	*.036*	* **.001**	**.19**	**.22**	*.064*	*.036*	* **.005**
GIVE CARE	**-.19**	**.11**				**.08**	**.07**	*.429*	*.492*	.30
FACTOR TOTAL	**-.29**	**-.17**	*.003*	*.049*	* **.001**	**.22**	**.31**	*.036*	*.003*	* **<.0005**

Table 5:51- Probabilities associated with Factor-based Difference Indices

For paradigmatic salience, its negative associations with the therapist/patient dissimilarity measure are shown to be reliable for all three general indices. Three out of the four composite subscales, presenting a more differentiated picture, are also reliably so associated; only the care-giving dimension fails to show the expected consistent relationship. The eight factor-based subscales provide an even more finely-grained view. Three of their correlations are reliable, each forming one half of one of the three composite scales showing significant associations.

The hypothesis, that paradigmatic mean salience scores will show negative associations with dissimilarity measures, is therefore fully supported for the general interpersonal distress indices. It is also supported for three composite subscales - two affiliative dimensions ('co-operate' and 'elicit care') and, most strongly, the dominance ('compete') dimension. It is further supported for three constituent factor-based subscales - 'too aggressive', 'hard to be social', and, most strongly, 'too dependent'. The hypothesis has not been supported for one of the composite subscales ('give care') and its components - 'too caring' and 'too open'. On this dimension, which might be taken to be prototypical for them in general, therapists do not perceive similarities between their own interpersonal problems and that of the patients which whom they experience paradigmatic difficulties. There are no noteworthy or reliable negative associations of transient salience with dissimilarity indices, in line with the predictions in the second half of the hypothesis.

As already observed, situational salience shows consistent positive associations with dissimilarity indices (except for one subscale - 'too caring'). This correlation is reliable only for one of the general interpersonal distress indices - the measure

calculated from the subset of the 32 factor-based items[44]. In addition, associations for three composite subscales (the same that featured in significant correlations with paradigmatic salience) are reliable, as are those for two of their constituent factor-based subscsales ('too aggressive' and 'hard to be social').

A second set of associations between difficulty type and measures of therapist / patient dissimilarity concerns the subset of 32 items forming the circumplex version of the IIP. Table 5:52 below shows the product-moment correlation coefficients for each circumplex octant and the aggregate general distress index, between mean salience scores for difficulty types and therapist/patient dissimilarities, both for untransformed and ipsatised scores, and for each of the two batches[45].

	ABSOLUTE DIFFERENCES						**IPSATISED ABSOLUTE DIFFERENCES**					
	PAR		***SIT***		TRANS		**PAR**		***SIT***		TRANS	
	B I	**BII**	***B I***	***BII***	B I	B II	**B I**	**BII**	***B I***	***BII***	B I	B II
DOMINEERG.	**-.32**	**-.15**	***.37***	***.25***	-.10	.14	**-.20**	**-.01**	***.25***	***.16***	.04	.08
INTRUSIVE	**-.03**	**.06**	***-.13***	***.21***	.13	-.09	**-.11**	**.05**	***.04***	***-.00***	.09	-.01
OVERLY NURTURANT	**-.05**	**.28**	***-.02***	***-.02***	.08	-.09	**-.25**	**.09**	***.25***	***.01***	.00	.08
EXPLOITABLE	**-.12**	**-.07**	***.03***	***.28***	-.06	-.08	**-.27**	**-.11**	***.20***	***.25***	.01	-.00
NON-ASSERTIVE	**-.20**	**-.19**	***.07***	***.22***	-.01	-.05	**.01**	**-.07**	***.12***	***.20***	-.09	-.17
SOCIALLY AVOIDANT	**-.18**	**-.14**	***.14***	***.24***	.01	.19	**-.07**	**.06**	***.04***	***.18***	05	.07
COLD	**-.20**	**-.06**	***.22***	***.11***	-.08	.13	**-.15**	**.00**	***.25***	***.01***	-.04	.12
VINDICTIVE	**-.32**	**-.05**	***.32***	***.12***	-.10	.21	**-.31**	**.08**	***.24***	***.24***	.01	-.02
CIRCUMPLEX TOTAL	**-.31**	**-.09**	***.22***	***.30***	-.02	.12	**-.29**	**.05**	***.32***	***.22***	.01	.07

Table 5:52 - Correlations between Difficulty Types and Circumplex Difference Indices

---

[44] The correlation for the positive item average (mean score for items endorsed higher than zero) would have been statistically significant but for the two-tailed test.

[45] In order to reduce skewness of their distributions, square root transformations were applied to Dominant, Intrusive, Overly Nurturant, Exploitable, Non-assertive, Vindictive.

Looking again for noteworthy associations among the untransformed scores in the left half of the table, none are to be found for correlations with transient salience (columns 5 and 6). In fact only three show the same sign in both batches. In contrast, associations with paradigmatic salience are consistently negative in all but two cases. Yet only two of the subscale correlations (those at either end of the dominance axis - Domineering and Non-assertive) meet the criterion of $r \geq |.15|$, though the Socially Avoidant octant comes close.

Turning to the associations with situational mean salience scores, all but two are positive. The total distress score calculated from the 32 circumplex items meets the above criterion, as does the Domineering subscale, while the Socially Avoidant and Vindictive octants come close.

As was true for factor-based measures, associations with ipsatised indices of dissimilarity are much less consistent and generally attenuated, indicating the importance of a substantive common distress factor in accounting for similarities.

Reliabilities of the observed associations are addressed in Table 5:53 below[46].

---

[46] As in Table 5:51, consistent correlations with paradigmatic and situational mean salience scores for each of the untransformed IIP indices, together with associated probabilities and combined probabilities are shown. In the absence of specific predictions, two-tailed tests have been used for situational salience. Combined probabilities meeting the significance criteria are displayed **bold** with asterisks.

	PARADIGMATIC MEAN SALIENCE SCORES					SITUATIONAL MEAN SALIENCE SCORES				
	r		*p*		p comb. (one-tail)	r		*p*		p comb. (two-tail)
	B I	**B II**	*B I*	*B II*		**B I**	**B II**	*B I*	*B II*	
CIRCUMPLEX TOTAL	**-.31**	**-.09**	*.001*	*.203*	* **.003**	**.22**	**.30**	*.035*	*.003*	* **<.0005**
DOMINEERING	**-.32**	**-.15**	*.001*	*.072*	* **.001**	***.37***	***.25***	*.001*	*.015*	* **<.0005**
INTRUSIVE	**-.03**	**.06**				***-.13***	***.21***			
OVERLY NURTURANT	**-.05**	**.28**				***-.02***	***-.02***			
EXPLOITABLE	**-.12**	**-.07**	*.122*	*.252*	.11	***.03***	***.28***	*.755*	*.007*	.030
NON-ASSERTIVE	**-.20**	**-.19**	*.025*	*.034*	* **.004**	***.07***	***.22***	*.493*	*.032*	.040
SOCIALLY AVOIDANT	**-.18**	**-.14**	*.042*	*.089*	.015	***.14***	***.24***	*.168*	*.021*	.009
COLD	**-.20**	**-.06**	*.030*	*.281*	.041	***.22***	***.11***	*.038*	*.302*	.028
VINDICTIVE	**-.32**	**-.05**	*.001*	*.320*	* **.005**	***.32***	***.12***	*.002*	*.244*	* **.002**

Table 5:53 - Probabilities associated with Circumplex Indices of Dissimilarity

For paradigmatic salience, its association with the therapist / patient dissimilarity measure is shown to be reliable for the general distress index (even though this did not meet the criterion for 'noteworthy' consistent magnitude). Reliable correlations are also shown for three of the octants - those at either pole of the dominance axis (Domineering and Non-assertive) and Vindictive (which describes hostile dominance). Although correlations for the Socially Avoidant and Cold subscales do not reach statistical significance, the overall picture is one of therapist / patient similarities in the left, hostile half of the interpersonal circle being associated with paradigmatic salience, while such associations in the right, affiliative half are absent or weak and unreliable.

The hypothesis, that paradigmatic mean salience scores will show negative associations with dissimilarity measures, is supported for the circumplex general distress index. It is also supported for three subscales, but not for the other five. There are no noteworthy or reliable negative associations of transient salience with dissimilarity indices, in line with the predictions in the second half of the hypothesis.

As already observed, situational salience shows consistent positive associations with dissimilarity indices, except for two subscales - Intrusive and Overly Nurturant. These correlations are reliable for the general distress index and for two adjacent octants - Domineering and Vindictive.

5.3.4.4 Summary of results and discussion

For the analyses in the preceding section, the IIP-51, a particular subset of items of the IIP, had been assembled. It allows for the calculation of indices according to two distinct structures - factor-based and circumplex-based - each corresponding to one of two validated short forms of the full inventory. The utility of this composite measure for tapping different dimensions of the domain of interpersonal distress has been demonstrated (for therapist self-report) through the relative segregation of subscales which theoretically should correspond to each other between each of the short forms. If this finding was to be replicated in the self-report of a patient sample, the IIP-51 might show promise as an economical and informative clinical measure.

For the current study, circumplex and factor-based forms provided discrete perspectives on the two main areas of interest: a) the prevalence of interpersonal problems as perceived by respondents in themselves and their patients, and b) the associations of the three difficulty types with measures of therapist / patient dissimilarity on a number of IIP indices. Prevalence data could be compared between

English- and German-language subsamples. The TDQ2 response format also allowed for comparisons between patient categories (those perceived as difficult and as not-so-difficult). Predictions had been made for the associations of similarity indices with transient and paradigmatic difficulty types. In the event, situational salience also yielded interpretable results. (Associations with difficulty type could only be investigated for the British sample, as salience ratings were not available for German-language accounts.)

The first consistent finding among the prevalence data was a monotone decline of interpersonal distress scores on most indices from difficult patients through not-so-difficult patients to therapists. With one exception (problems with excessive openness), therapists as a group view themselves as least distressed. With two exceptions (problems with assertiveness and excessive caring), they view their difficult patient as most distressed, while their not-so-difficult patients are on the whole perceived as lying in the middle. These findings are constant across both language-samples. As already observed above, difficult patients are patients with interpersonal difficulties.

Therapists perceive themselves and their patients in either category as differentiated not only by the level but also by the nature of their interpersonal problems. Through the prism of factor-based measures, therapists see themselves as too open and caring and not aggressive and assertive enough; through that of circumplex measures they appear as excessively nurturant, non-assertive and somewhat gullible. These findings are quite consistent across the two cultures (though German therapists are perhaps less conflicted about their aggression and assertiveness). If one reframes therapists' interpersonal problems positively, they appear as undefended, honest, caring, friendly,

prepared to be led, and ready to negotiate; in short, prototypically Rogerian. Their assets in the consulting room appear as liabilities outside it.

In factor-based terms, both difficult and not-so-difficult patients are viewed by their therapists as lacking in openness and capacity for co-operation; deficits which would make it hard for them to form and sustain a working alliance. In addition, difficult patients are seen as uncaring and unsupportive, while not-so-difficult patients are viewed as unassertive and unable to challenge. Circumplex measures replicate these perceptions - the interpersonal problems of difficult patients are placed in the left, hostile half of the interpersonal circle, while those of not-so-difficult patients are located in the bottom, sub-assertive half. Again, there is fair consistency across language-subsamples.

One further interesting difference emerges from the comparisons with published samples. Dependency problems, prevalent amongst an out-patient sample, are characteristic of neither patient category in the current samples. While this finding is weakened by the TDQ2 data being based on therapist ratings, whereas the comparison sample is based on patient self-report, it does, however, tally with an impression gathered by reading the difficulty accounts. There is a remarkable absence of narratives of therapists' difficulties in extricating themselves from long-term therapeutic relationships with dependent patients who are neither improving nor letting go. Yet in supervision or discussions among peers, these seem to feature regularly. One might speculate that such difficulties feel ego-syntonic enough for therapists not to mention them when asked to recall specific difficult situations, but are burdensome enough, to feature in day-to-day discourse.

Associations of transient and paradigmatic difficulty types with measures of therapist / patient dissimilarity over a variety of IIP indices are largely in line with the

hypothesis: consistently and mostly reliably negative for paradigmatic difficulties and absent for transient difficulties. Exceptions are to be found in the factor-based care-giving dimension and in the affiliative half of the circumplex. With patients with whom they experience paradigmatic difficulties, therapists perceive similarities in those areas of interpersonal distress, which are generally uncharacteristic for them. It may be that these are the broad areas they are conflicted in, having adapted to those conflicts by their choice of profession. Alternatively, they may have adapted their perceptions of the patients in question, seeing them as less distressed and therefore more similar to themselves, to reduce conflict. Even though the former explanation seems more plausible, the link between paradigmatic difficulties and therapists' internal conflicts is supported either way, lending substance to both the paradigmatic and the transient difficulty concept.

Associations of situational salience both with IIP indices and with measures of therapist / patient dissimilarity furnished somewhat less expected findings. Positive associations with IIP subscales and composite indices were not in themselves surprising, given the association of situational salience with 'difficult' patient category (see 5.3.2.1 above) which is positively correlated with interpersonal distress. What is remarkable though, is that positive associations for situational salience are frequently the 'mirror image' of negative ones for paradigmatic salience, showing substance and/or significance on the same indices but in opposite directions. This is particularly true for associations with dissimilarity measures. Patients who cause situational type difficulties are seen as 'other' by therapists on precisely those measures on which they characterise themselves as a group. This finding is conceptually consistent with the definition of situational difficulties as likely to be experienced as such by most therapists, regardless of personal attributes and level of

competency; and therefore contributes to the validation of the situational difficulty type.

5.3.5 SASB INTREX Data

The INTREX questionnaire has been described above in 5.2.1.1.6. The version used in TDQ2 provides information about interpersonal transactions during the difficulty, but also of therapists' corresponding ('situation-specific') and generalised internal states ('baseline'). It further allows for the third general hypothesis (see 5.1.2 above) to be operationalised and tested. Software for analysing SASB data (see 5.2.5 above) generates a comprehensive array of indices and coefficients. If all options are used, the ratio of output to input variables is approximately 3 : 1![47] Data analysis therefore has to be selective and consequently only a small number of the possible indices have been used in the current study. Apart from cluster (subscale) scores, which are derived from single item ratings in the INTREX short form, two composite types of measures have been used - pattern coefficients (attack, control, and conflict) and weighted scores (affiliation and autonomy).

Pattern coefficients describe the degree of fit between observed distributions of cluster scores and theoretical curves. Positive attack coefficients describe patterns organised around the hostile pole of the affiliation axis of each surface, negative attack coefficients those organised around the friendly pole. Positive control coefficients describe patterns organised around enmeshment, negative control coefficients those around autonomy. Conflict coefficients describe patterns showing high endorsements at opposite poles - positive values indicate autonomy conflicts, negative values affiliation conflicts (Benjamin, 1984). Pattern coefficients tend to be

---

[47] For TDQ2, this means that well in excess of 50,000 data points would be available for further analysis.

bi-modally distributed (Benjamin, 1995) and are probably most useful for within-subject analyses.

Weighted scores represent respondents' positions on the circular surface projected onto the axes - horizontal for affiliation scores and vertical for autonomy scores - corrected for angular distortion. They tend to be normally distributed (Davies-Osterkamp et al., 1993) and are useful for between-subjects analyses.

As in the analysis of IIP data, INTREX indices have been correlated first with patient categories and subsequently with mean salience ratings for difficulty types. Analyses reported in this section start with therapist introjects and progress to therapist / patient transactions. Finally, the first general hypothesis outlined in 5.1.2 above is operationalised for SASB INTREX indices and tested.

5.3.5.1 INTREX Data and Patient Categories

As for IIP data, the first step in the analysis of SASB ratings consists of investigating how the therapists responding to TDQ2 saw themselves and their difficult and not-so-difficult patients.

5.3.5.1.1 Therapist Introjects

SASB INTREX surface 3, the rater's introject, is presented four times in TDQ2. Two of these presentations are situation-specific, measuring therapist introjects at the time when the difficulty occurred, while the other two are 'baseline' introjects, measuring how respondents generally treat themselves 'at best' and 'at worst'. Descriptive statistics from all four presentations of the eight items pertaining to surface 3 are shown in Table 5:54 below.

	THERAPIST INTROJECT							
	ENGLISH DATA SET N = 92				GERMAN DATA SET N = 27			
	At Best	***At Worst***	With NSD	*With DP*	**At Best**	***At Worst***	With NSD	*With DP*
Spontaneous self	**55.9 (24.7)**	***31.8 (22.8)***	32.8 (26.2)	*25.1 (21.4)*	**76.1 (18.4)**	***34.8 (19.7)***	42.6 (24.1)	*38.8 (27.4)*
Self-accepting and exploring	**72.9 (18.7)**	***33.1 (20.1)***	52.7 (23.5)	*48.4 (23.4)*	**76.8 (20.8)**	***34.4 (26.2)***	64.0 (23.6)	*52.2 (26.3)*
Self-loving and cherishing	**62.7 (23.4)**	***26.6 (19.7)***	41.1 (23.5)	*36.8 (21.8)*	**73.6 (16.7)**	***33.3 (25.6)***	44.7 (29.0)	*40.0 (28.7)*
Self-nourishing and enhancing	**70.3 (18.7)**	***32.3 (21.6)***	49.7 (23.5)	*47.3 (22.9)*	**71.4 (19.0)**	***43.7 (20.7)***	55.5 (24.7)	*47.3 (26.9)*
Self-monitoring and restraining	**36.7 (21.8)**	***56.6 (27.0)***	44.0 (27.7)	*53.0 (22.1)*	**35.2 (21.7)**	***55.9 (22.6)***	57.7 (30.3)	*60.8 (24.5)*
Self-indicting and oppressing	**21.5 (18.6)**	***54.9 (27.7)***	33.4 (26.7)	*39.9 (25.2)*	**8.9 (14.2)**	***57.3 (31.8)***	25.9 (28.2)	*46.5 (28.2)*
Self-rejecting and destroying	**12.6 (15.1)**	***50.8 (29.6)***	16.3 (23.0)	*21.8 (23.2)*	**4.1 (7.5)**	***28.9 (30.5)***	5.6 (11.2)	*19.2 (22.6)*
Daydreaming and neglecting of self	**26.7 (26.1)**	***36.4 (27.9)***	14.7 (21.1)	*11.7 (15.5)*	**11.5 (16.8)**	***34.8 (23.6)***	14.4 (18.5)	*25.8 (23.2)*
Weighted affiliation score	**102.3 (55.1)**	***-38.1 (68.7)***	55.3 (69.0)	*39.4 (66.1)*	**139.7 (45.1)**	***-3.5 (76.0)***	82.7 (63.9)	*35.6 (82.3)*
Weighted autonomy score	**23.1 (44.4)**	***-34.1 (52.1)***	-18.5 (52.0)	*-42.5 (38.9)*	**44.4 (38.0)**	***-38.7 (44.8)***	-16.4 (46.9)	*-30.3 (51.3)*
% Positive attack coefficient	**3.1 [0]**	***74.5 [20.4]***	22.8 [4.3]	*40.2 [0]*	**0 [0]**	***59.3 [18.5]***	7.4 [0]	*46.2 [7.7]*
% Positive control coefficient	**68.3 [5.1]**	***60.2 [4.1]***	81.5 [14.1]	*93.5 [26.1]*	**70.4 [0]**	***81.5 [7.4]***	88.9 [3.7]	*80.8 [7.7]*
% Positive conflict coefficient	**66.3**	***62.2***	75.0	*79.3*	**55.6**	***81.5***	85.2	*84.6*
Mean attack coefficient	**-.75 (.28)**	***.36 (.57)***	-.40 (.57)	*-.21 (.62)*	**-.85 (.12)**	***.17 (.63)***	-.55 (.37)	*-.11 (.63)*
Mean control coefficient	**.22 (.50)**	***.15 (.53)***	.44 (.44)	*.58 (.32)*	**.22 (.46)**	***.38 (.44)***	.43 (.36)	*.40 (.44)*

Table 5:54 - SASB INTREX, Therapist Introject: Descriptive Statistics

The first eight rows show mean scores (with standard deviations in brackets) for the eight clusters (octants) of the SASB Introject surface, starting at the extreme autonomy pole and moving clockwise around the circumplex. (Cluster labels are shown in the left margin.) Mean weighted affiliation and autonomy scores are shown in the next two rows. As attack, control and conflict coefficients generally tend to

have bimodal distributions either side of zero, they are not useful for means comparisons. The next three rows therefore show the percentage of values above zero occurring for each of the coefficients. For attack and control coefficients, figures in square brackets give the percentages of coefficients ≥ .82, which is the threshold for individual significance ($p \leq .05$). Means for attack and control coefficients are, however, shown in the last two rows as they permit comparison with some normative samples (see below). Columns show baseline values for therapist introject (in private relationships) at best and at worst, followed by the specific introjects during difficulties with 'not so difficult' and 'difficult' patients, for both British and German-language data sets.

Looking down columns, it is apparent that cluster mean scores by and large follow a circular pattern with monotone rises and falls either side of the extreme values. Therapist introject 'at best' is characterised by high mean endorsements on the first four clusters ('give autonomy' to 'friendly control') in both data sets with peaks occurring on cluster 2 ('give friendly autonomy'). For therapist introject 'at worst', high mean endorsements occur at clusters 5 and 6 ('self-control', 'hostile self-control'); in the British data set the extreme disaffiliation pole ('self-rejecting and destroying') is also highly rated, quite in contrast to the German-language sample. Situation-specific introjects are also highly endorsed on the first four clusters, those with 'difficult' patients show a second peak on cluster 5 (self-control).

Comparing mean scores across columns by rows, a pattern is discernible of 'at best' and 'at worst' cluster means defining minima and maxima of a range, while the values for situation-specific introjects lie between the two. Eight-point correlation coefficients, calculated from mean cluster scores, show a close correspondence

between values for 'not so difficult' patients and 'at best' baseline in both data sets (r = .81 / r = .80), for 'difficult' patients these associations are smaller (r = .50 / r = .45). Baseline introjects are, not surprisingly, inversely related - more clearly so in the British sample (r = -.77 / r = -.23). These findings are consistent with a view that a therapist's 'working self', as expressed by the SASB introject, lies somewhere between the extreme self-representations associated with 'best' and 'worst' states; but that difficult patients give rise to more negative introjects in their therapists than not-so-difficult patients. It is, however, worth bearing the general point in mind, that all cluster means have to be treated with some caution as they rely on single item ratings.

Weighted affiliation and autonomy scores, on the other hand, are each calculated from six cluster scores and lend themselves to means comparisons. The difference between mean weighted affiliation scores for therapist introjects with difficult and not-so-difficult patients is statistically significant, both for the British sample (t = 2.08, df = 89, p = .041) and for the German-language sample (t = 3.45, df = 24, p = .002. Combined probability: z = 5.24, p < .0001). The corresponding difference between mean weighted autonomy scores is also statistically significant for the British sample (t = 4.91, df = 89, p < .0001), but not for the German-language sample (t = 1.29, df = 24, p = .209). Treating one of the results as a replication, however, the significance of the difference is clearly established (combined probability: z = 5.57, p < .0001). Baseline introjects, it appears, differ sharply in terms of friendliness vs. hostility directed against the self in best and worst states. Difficulty-specific introjects show moderately positive affiliation; those engendered by not-so-difficult patients are significantly friendlier and therefore closer to the 'at best' introject than those evoked by difficult patients. In terms of autonomy vs. control, difficulty-specific introjects

are characterised by high self-restraint. Those engendered by difficult patients are closer to (and, in the British sample, in excess of) baseline introjects 'at worst'.

Percentages of positive attack and conflict coefficients generally confirm this picture, although the relationship between self-control and patient category is equivocal. Conflict coefficients are predominantly positive, indicating a preponderance of conflicts on the control dimension (co-presence of high 'spontaneous self' and 'self-monitoring and restraining' endorsements), across all subsamples.

Normative data for baseline introjects are not easily available. Despite the problem of bi-modality, Benjamin (1995) quotes coefficient means for a sample of 500 college students. Mean attack coefficients for introject 'at best' were -.648 / -.617, those for introject 'at worst' .243 / -.018, depending on the Short Form version used.[48] In a German validation study, Davies-Osterkamp et al. (1993) report 'at best' mean attack coefficients of -.82 for a student sample (n = 101) and -.49 for a patient sample (n = 51), the corresponding values 'at worst' being -.11 and .50. Comparing these results with the attack coefficient means reported in Table 5:54, it appears that 'at best' introjects of respondents to TDQ2 are very similar to student samples, whereas values for 'at worst' introjects lie in between those of the student samples and that of the patient sample, a result suggestive of a 'wounded healer' view of psychotherapists. Mean control coefficients in Benjamin's student sample were .496 / .515 'at best' and .042 / .151 'at worst', again depending on the Short Form version used. Corresponding values from the German study are .54 for students and .33 for patients 'at best', and .40 for students compared with .03 for the patient sample 'at

[48] In another student sample, for which N is not reported, mean attack coefficients of -.80 'at best' and .06 'at worst' were found (Benjamin, 1984).

worst'. Values for TDQ2 respondents are in a similar range but show an equivocal picture.

Published normative data for mean weighted affiliation and autonomy scores are, to the best of my knowledge, confined to the German study cited above. Table 5:55 below summarises those results in comparison with the results obtained with TDQ2 (Standard Deviations in brackets):

	Therapists TDQ 2				**German Comparison Groups**			
	British Data (N = 92)		**German Data** (N = 27)		**Students** (N = 101)		**Patients** (N = 51)	
	best	worst	best	worst	best	worst	best	worst
Weighted Affiliation Score	**102.3** **(55.1)**	***-38.1*** ***(68.7)***	**139.7** **(45.1)**	***-3.5*** ***(76.0)***	**152.4** **(50.2)**	***21.7*** ***(93.5)***	**90.0** **(91.4)**	***-82.6*** ***(89.1)***
Weighted Autonomy Score	**23.1** **(44.4)**	***-34.1*** ***(52.1)***	**44.4** **(38.0)**	***-38.7*** ***(44.8)***	**0.2** **(47.4)**	***-39.1*** ***(55.5)***	**9.9** **(46.1)**	***-36.6*** ***(66.0)***

Table 5:55 - Weighted Scores: TDQ Samples Compared with Normative German Samples

Therapists' mean affiliation scores lie between the extremes defined by the student sample ('at best') and the patient sample ('at worst'). The high mean for students may perhaps be explained by the notion that friendly overinvestment in self is a usual (and probably necessary) characteristic of 'prolonged adolescence' (Blos, 1962). Mean autonomy scores 'at worst' are remarkably consistent across all four samples. Therapists show, however, a much greater propensity towards treating themselves in a spontaneous and unguarded manner 'at best', but not when in role (see Table 5:54).

In summary, therapists' introjects, as reported in response to TDQ2, range widely between 'best' and 'worst' states, indicating a broad spectrum of self-representations in these samples. Positive affiliation 'at best' is commensurate with reported comparison groups, both clinical and non-clinical. Negative affiliation 'at worst' is more marked than in student comparison groups but less severe than in a reported

patient sample, suggesting that therapists are either more aware of or more prone to self-attacks than a non-clinical population.[49] Difficulty-specific affiliation is on average mildly positive, indicating that therapists by and large manage to stay on good terms with themselves when challenged professionally, though they are more tested in this respect by patients whom they regard as 'difficult'.

For the autonomy dimension, therapists' baseline introjects are characterised by high positive values in the 'best' state, especially when combined with positive affiliation (see cluster 2 scores), indicating that they strongly adopt the classical Rogerian position of 'accepting and exploring' towards themselves - much more so than respondents from non-clinical and clinical comparison groups. Difficulty-specific introjects are strongly characterised by 'self-monitoring and restraint', suggesting that therapists' spontaneity is sharply curtailed when they are professionally challenged, causing them to adopt a much more restrictive (and probably defensive) stance towards themselves. This would, however, appear an entirely appropriate coping strategy, as long as it was combined with positive affiliation.

5.3.5.1.2 Therapist-Patient Transactions

Alongside introjects, the other two surfaces measured by SASB INTREX are 'transitive' (referent active) and 'intransitive' (referent reactive). Respondents to TDQ2 rated each of these surfaces for each difficulty, both with the therapist as referent and with (the therapist's perception of) the patient as referent.

---

[49] Of course, levels of affiliation in a particular state do not provide information about the persistence of that state - a consideration which led Essex, Klein, Lohr, & Benjamin, (1985) to ask their respondents how much time they spent in each state.

Table 5:56 - INTREX Transactions British Sample	Transitive (Active) Surface 1				Intransitive (Reactive) Surface 2				N = 92
	Therapist with Patient		Patient with Therapist		Therapist with Patient		Patient with Therapist		
	DP	*NSD*	DP	*NSD*	DP	*NSD*	DP	*NSD*	
Freeing and Forgetting	**31.3** (23.1)	*36.3* *(25.0)*	**18.9** (22.9)	*32.5* *(23.1)*	**50.9** (22.2)	*49.2* *(25.2)*	**46.8** (30.9)	*45.6* *(27.3)*	**Asserting and Separating**
Affirming and Understanding	**50.5** (24.0)	*65.7* *(20.1)*	**30.8** (24.2)	*48.3* *(24.8)*	**55.5** (24.4)	*59.0* *(25.2)*	**26.1** (27.3)	*39.1* *(25.2)*	**Disclosing and Expressing**
Loving and Approaching	**39.2** (25.7)	*49.8* *(24.1)*	**12.2** (16.8)	*21.5* *(20.3)*	**15.2** (17.9)	*34.1* *(25.5)*	**14.8** (20.9)	*28.9* *(26.0)*	**Joyfully Connecting**
Nurturing and Protecting	**42.0** (23.7)	*49.8* *(26.4)*	**14.4** (19.0)	*22.6* *(20.4)*	**17.9** (20.1)	*29.6* *(25.6)*	**26.6** (25.3)	*47.9* *(26.0)*	**Trusting and Relying**
Watching and Controlling	**33.9** (26.9)	*27.4* *(27.0)*	**26.4** (28.8)	*16.2* *(18.2)*	**26.6** (22.9)	*28.0* *(26.1)*	**24.7** (23.1)	*27.1* *(22.2)*	**Deferring and Submitting**
Belittling and Blaming	**24.2** (25.8)	*16.3* *(22.1)*	**51.9** (33.6)	*27.8* *(31.6)*	**18.3** (19.9)	*14.7* *(21.9)*	**22.7** (23.9)	*13.8* *(16.6)*	**Sulking and Scurrying**
Attacking and Rejecting	**7.0** (10.5)	*4.8* *(9.4)*	**39.9** (32.5)	*19.3* *(28.7)*	**21.4** (24.4)	*13.7* *(23.8)*	**33.0** (31.2)	*22.9* *(27.6)*	**Protesting and Recoiling**
Ignoring and Neglecting	**15.8** (22.1)	*10.8* *(18.7)*	**33.8** (28.4)	*23.9* *(26.8)*	**23.3** (22.8)	*21.2* *(23.0)*	**53.3** (29.9)	*41.0* *(27.5)*	**Walling off and Distancing**
Weighted Affiliation Score	**61.0** (58.2)	*94.1* *(52.1)*	**-49.8** (76.7)	*12.9* *(75.6)*	**11.93** (57.0)	*49.6* *(61.3)*	**-30.9** (64.1)	*23.8* *(72.2)*	**Weighted Affiliation Score**
Weighted Autonomy Score	**-2.47** (41.2)	*14.9* *(44.1)*	**-8.21** (52.42)	*28.1* *(37.5)*	**47.6** (41.6)	*40.3* *(47.2)*	**38.0** (61.3)	*28.4* *(54.5)*	**Weighted Autonomy Score**
% Positive Attack Coefficient	**13.0** [1.1]	*5.4* *[O]*	**67.4** [20.7]	*35.9* *[9.8]*	**25.0** [2.2]	*9.2* *[1.1]*	**62.0** [8.7]	*28.3* *[2.2]*	**% Positive Attack Coefficient**
% Positive Control Coefficient	**75.0** [4.3]	*72.8* *[6.5]*	**35.9** [1.1]	*22.8* *[0]*	**15.3** [0]	*29.3* *[0]*	**23.9** [0]	*42.4* *[4.4]*	**% Positive Control Coefficient**

Table 5:57 -INTREX Transactions German Sample	Transitive (Active) Surface 1				Intransitive (Reactive) Surface 2				N = 27
	Therapist with Patient		Patient with Therapist		Therapist with Patient		Patient with Therapist		
	DP	*NSD*	DP	*NSD*	DP	*NSD*	DP	*NSD*	
Freeing and Forgetting	**36.2** (26.8)	*60.0* *(28.0)*	**31.5** (33.5)	*45.4* *(35.8)*	**40.4** (27.5)	*50.3* *(26.9)*	**52.7** (25.7)	*52.5* *(26.5)*	**Asserting and Separating**
Affirming and Understanding	**50.8** (20.2)	*64.4* *(25.9)*	**21.9** (18.3)	*41.9* *(26.9)*	**49.2** (30.6)	*57.3* *(24.9)*	**41.9** (28.0)	*64.0* *(26.0)*	**Disclosing and Expressing**
Loving and Approaching	**38.1** (24.3)	*51.5* *(28.4)*	**10.0** (12.7)	*30.7* *(28.2)*	**17.3** (17.6)	*32.2* *(25.3)*	**20.0** (27.3)	*42.9* *(25.8)*	**Joyfully Connecting**
Nurturing and Protecting	**47.3** (23.3)	*65.5* *(19.4)*	**6.5** (8.5)	*21.5* *(22.1)*	**42.7** (28.8)	*48.9* *(31.7)*	**32.3** (22.7)	*60.3* *(25.5)*	**Trusting and Relying**
Watching and Controlling	**45.0** (23.0)	*30.4* *(23.9)*	**32.3** (29.7)	*24.8* *(25.3)*	**25.8** (23.2)	*25.6* *(22.1)*	**14.6** (14.2)	*18.5* *(20.1)*	**Deferring and Submitting**
Belittling and Blaming	**19.6** (21.1)	*5.9* *(14.7)*	**47.5** (36.6)	*17.7* *(29.4)*	**18.3** (19.9)	*13.3* *(19.0)*	**31.9** (26.2)	*11.5* *(18.1)*	**Sulking and Scurrying**
Attacking and Rejecting	**6.9** (20.4)	*0.7* *(3.9)*	**31.1** (33.8)	*11.1* *(25.2)*	**21.6** (18.0)	*2.2* *(6.4)*	**25.0** (26.1)	*10.3* *(25.1)*	**Protesting and Recoiling**
Ignoring and Neglecting	**6.9** (11.6)	*4.8* *(11.2)*	**34.6** (26.3)	*17.7* *(28.1)*	**13.9** (19.6)	*11.1* *(20.4)*	**36.9** (30.0)	*20.0* *(26.8)*	**Walling off and Distancing**
Weighted Affiliation Score	**70.6** (61.7)	*116.5* *(46.8)*	**-50.8** (70.3)	*33.6* *(81.6)*	**27.9** (49.6)	*75.2* *(51.6)*	**-1.82** (61.5)	*84.1* *(67.7)*	**Weighted Affiliation Score**
Weighted Autonomy Score	**-13.8** (49.4)	*27.6* *(42.9)*	**0.56** (51.7)	*30.5* *(41.4)*	**20.6** (55.1)	*27.7* *(48.8)*	**45.3** (39.2)	*40.0* *(59.7)*	**Weighted Autonomy Score**
% Positive Attack Coefficient	**15.4** [3.8]	***0.0*** *[O]*	**69.2** [26.9]	***14.8*** *[3.8]*	**26.9** [0]	***0.0*** *[0]*	**38.5** [0]	***11.1*** *[0]*	**% Positive Attack Coefficient**
% Positive Control Coefficient	**84.6** [26.9]	***74.1*** *[0]*	**34.6** [0]	***37.0*** *[0]*	**42.3** [7.7]	***55.6*** *[0]*	**19.2** [0]	***55.6*** *[3.7]*	**% Positive Control Coefficient**

5.3.5.1.2.1 Prevalence

Combined descriptive statistics for surface 1 ('transitive') and surface 2 ('intransitive') are displayed above in Table 5:56 for the British sample and in Table 5:57 for the German-language sample. As in Table 5:54, figures in the first eight rows show means (and standard deviations in brackets) for cluster scores in clockwise order starting with cluster 1, the topmost octant, in the first row. Cluster labels for surface 1 are displayed in the left, those for surface 2 are shown in the right margin. Mean weighted affiliation and autonomy scores are shown in the next two rows, followed by percentages for positive attack and control coefficients [together with percentages of coefficients ≥ .82 in square brackets]. Conflict coefficients have been omitted, as they tend to be largely positive[50] and contribute little to the issues under investigation here. This form of display allows for the comparison of a number of relationships:

a) Similarities between patient categories are shown in adjacent columns (**DP/*NSD***) for each referent on each surface;

b) Interpersonal Complementarities, defined as identical topological positions adopted by two referents on two different surfaces, are indicated by comparisons across rows of columns with the same patient category (**DP or *NSD***), but with different referents on the two surfaces (for instance column pairings 1&7 or 4&6);

c) Interpersonal Similarities, defined as identical topological positions adopted by two referents on matching surfaces, are indicated by comparisons across rows of columns with the same patient category (**DP or *NSD***), but with different referents on the surface in question (for instance column pairings 1&3 or 6&8).

---

[50] A finding also confirmed by Tress & Junkert (1993).

Interpersonal Antitheses, defined as diametrically opposite topological positions adopted by two referents on two different surfaces, are not easily discernible from the table. They involve comparing the same columns as Interpersonal Complementarities but shifting by four rows between referents (for instance pairings of column 1 / row 1 with column 7 / row 5, or column 4 / row 7 with column 6 / row 3).

Negative Complementarities (Henry, Schacht, & Strupp, 1990) are simply Interpersonal Complementarities confined to negative affiliation clusters (rows 6 - 8).

Statistical estimates of relationships a) - c) can be made by correlating relevant pairs of scores across the eight clusters, or comparing the relevant pair of mean weighted scores.

Before turning to these relationships, some observations can be made by inspecting the patterning of ratings down the columns, which should theoretically show the monotone rise and fall of circular ordering. By and large this is the case, with the exceptions noted below.

As one would expect, therapists' highest scores occur in the second octant (friendly autonomy) consistently on both surfaces for both patient categories in both data sets. For German respondents, there is a second peak in the fourth octant (friendly enmeshment). Lowest scores are equally predictably organised around the extreme disaffiliative pole. This picture is also borne out by the weighted affiliation scores, which are consistently positive (but markedly higher in respect of not-so-difficult patients), and by the equally consistent low percentages of positive attack coefficients. Weighted autonomy scores are generally positive, the exception being therapists' active transactions with difficult patients which are just on the negative (controlling) side in the English data set and a little more pronounced in the German sample.

Difficult patients are experienced as actively hostile, with a peak occurring in octant 6 (hostile control) in both German and English data. Their reactions are organised around the autonomy pole, with high scores occurring in the first octant for both samples. For the British sample another high score falls in octant 8 (hostile autonomy), whereas for the German sample it occurs, somewhat surprisingly, in octant 2 (friendly autonomy). Not-so-difficult patients show the same autonomous reaction - more friendly in the German sample, more hostile in the British sample - but are also perceived as trusting and relying (octant 4). Their actions are also organised around the autonomy pole with a bias towards the friendly side (octant 2).

This representation of the interpersonal transactions of patients (as their therapists see them) during the occurrence of difficulties can now be compared to the picture of their more enduring interpersonal problems provided by the circumplex version of the IIP, as outlined in the previous section (5.3.4.1.2). Difficult patients appear to display the same disaffiliative interpersonal behaviours during difficulty situations that characterise their predominant long-term interpersonal problems. (In the German sample, this occurs alongside some 'friendly autonomy' interactions, which is somewhat counterintuitive and hard to interpret.) It might be argued generally that the correspondence between situation-specific and long-term interaction patterns is an artefact, because respondents might have generalised from their patients' behaviour during difficulties when assessing their enduring problems. The results for not-so-difficult patients, however, provide some reassurance that respondents were able to differentiate between the two conditions. Patients in this category are characterised by long-term problems organised around the negative pole of the autonomy/dominance dimension. Their difficulty-specific actions and reactions, in sharp contrast, are organised around the positive pole of the same

dimension. Not-so-difficult patients, on average, behave quite out of character during difficulty situations, which might well be one of the underlying reasons why they are experienced by their therapists as deviating from their customary easy presentation. Further clarification of this issue could be sought by within-subject comparisons; however, such an investigation would be outside the scope of this thesis.

Having examined scores within columns of Table 5:56 and Table 5:57, we can now return to the comparisons between columns described in a) - c) above.

5.3.5.1.2.2 Direct Comparisons between Patient Categories

The first of these comparisons concerns the similarities between patient categories expressed through mean cluster scores (and associated measures) for each referent for each of surfaces 1 and 2. Correlation coefficients (calculated from the eight mean cluster scores) for the associations between patient categories for both data sets are shown in Table 5:58 below:

	CORRELATIONS BETWEEN PATIENT CATEGORIES			
	SURFACE 1		**SURFACE 2**	
	Therapist with Patient	Patient with Therapist	Therapist with Patient	Patient with Therapist
ENGLISH DATA SET	**.96**	**.09**	**.85**	**.49**
GERMAN DATA SET	**.89**	**-.25**	**.88**	**.50**

Table 5:58 - Therapist/Patient Transactions: Correlations between Patient Categories

In both data sets therapists perceive the patterns of their interactions with difficult and not-so-difficult patients to be very similar - as indicated by correlation coefficients of .85 and above - both in active (surface 1) and reactive (surface 2) mode. In sharp contrast, transitive interactions by patients in the two categories are seen as quite

dissimilar; either as near orthogonal, in the British data set, or in the form of a weak inverse relationship, in the German data set. Intransitive interactions by patients in both categories, however, are perceived as being moderately positively associated with each other.

While these data do not give any indication about temporal sequences, the results are, nevertheless, quite consistent with the following speculative view. In difficult situations, therapists, on average, approach their patients in much the same way, regardless of whether they find them 'difficult' or 'not-so-difficult'. The approach they experience from patients in the two categories are, however, very different. In response, therapists do manage to treat both categories of patients in a similar fashion, which results in a convergence between the reactive states of patient in either category. A less speculative account might be that therapists on average manage to be consistent in their approach to patients and that their evaluation of patients as being either difficult or not-so-difficult is based more on how they are approached rather than how they are reacted to. Clinical experience, however, would favour the former view.

5.3.5.1.2.3 Therapist- Patient Complementarities

High degrees of complementarity within the SASB system are regarded as indicators of stability within relationships (Benjamin, 1984). Despite the possibility of stable destructive patterns (the 'negative complementarity' Henry et al. (1990) describe), which should be comparatively rare, one would generally expect high complementarity to be a sign of a collaborative set between therapist and patient and therefore as an indication of a functional working alliance. Table 5:59 below shows correlation coefficients (based on the eight pairs of mean cluster scores) between complementary surfaces for each data set. Figures in the first two columns show complementarities -

differentiated by patient category - for transitive therapist and intransitive patient surfaces (columns 1 & 7 and columns 2 & 8 of Table 5:56 and Table 5:57). Figures in the last two columns represent the corresponding complementarities for transitive patient and intransitive therapist surfaces (columns 3 & 5 and 4 & 6).

	COMPLEMENTARITIES			
	Therapist active -Patient reactive		Patient active - Therapist reactive	
	DP	NSD	DP	NSD
ENGLISH DATA SET	**-.48**	**.49**	**-.09**	**.77**
GERMAN DATA SET	**.02**	**.96**	**-.41**	**.84**

Table 5:59 - Therapist/Patient Transactions: Correlations between Complementary Surfaces

There is a clear division between patient categories apparent from the results with complementarities high for English not-so-difficult patients and even higher for German ones; complementarities are conversely low (even clearly antithetical in two instances) for difficult patients in both samples. Interactional patterns with difficult patients are thus perceived as highly volatile, whereas those with not-so-difficult patients are perceived as complementary, stable and, by implication, more positive.

5.3.5.1.2.4 Interpersonal Similarities

Interpersonal similarities within the SASB system have been variously interpreted as evidence of identification (Benjamin, 1988a), imitation (Benjamin, 1994b), introjection (Benjamin, 1996b), or interpersonal modelling (Benjamin, 1995). As such, they should be closely related to the three 'copy processes' of early childhood experience - identification/imitation for surface 1, internalisation for surface 2, and introjection for surface 3 (Henry, 1996). These relationships are, however, so far not elaborated in the literature. In the current context, interpersonal similarities are best understood as indicators of how much therapists perceived their patients' transactions as being like their own in the course of the difficulties they experienced with them. Table 5:60 below shows correlation coefficients between the mean cluster scores of therapists and patients for each patient category on each surface for each data set.. Figures in the first two columns show similarities for the transitive surface (columns 1 & 3 and columns 2 & 4 of Table 5:56 and Table 5:57). Figures in the last two columns represent the corresponding similarities for the intransitive surface (columns 5 & 7 and 6 & 8).

	INTERPERSONAL SIMILARITIES			
	SURFACE 1		**SURFACE 2**	
	DP	NSD	DP	NSD
ENGLISH DATA SET	**-.57**	**.58**	**.32**	**.59**
GERMAN DATA SET	**-.58**	**.79**	**.54**	**.96**

Table 5:60 - INTREX: Therapist / Patient Interpersonal Similarities

The results show a clear perceived dissimilarity between therapists and their difficult patients as far as interpersonal actions are concerned. Reactions, however, are seen as moderately to substantially similar. Not-so-difficult patients are perceived as quite

similar to their therapists in their difficulty-specific transactions, up to the point where their reactions are evaluated as practically identical to those of their therapists in the German sample. These findings are in part consistent with those from IIP indices described in 5.3.4.3 above. Therapists' have similar perceptions of their patients' situation-specific interpersonal actions and general interpersonal problems - difficult patients are unlike themselves, not-so-difficult patients more like themselves. Reactions, however, are far less differentiated by patient category, indicating that respondents to TDQ2 did not simply extrapolate from difficulty-specific to general interpersonal behaviours of their difficult patients.

5.3.5.2 INTREX Data and Difficulty Types

As for IIP data, there are two ways of analysing the relationships of SASB INTREX data and difficulty types. Mean salience scores can either be correlated directly with INTREX scores or they can be analysed in relation to a similarity or distance measure.

5.3.5.2.1 Direct Associations

If there are any associations of difficulty types with indices derived from SASB INTREX, one would expect these to provide information about underlying processes and interpersonal substrates of individual difficulty experiences. Correlations with cluster scores are likely to provide supplementary information but need to be treated with caution as they rely on single items. attack and control coefficients, as noted above, are unsuitable for correlational analysis because of their likely bi-modal distribution. Weighted affiliation and autonomy Scores appear the most suitable variables for analysis. Affiliation scores are of particular interest because of the likely relationship of hostility to problematic interpersonal process and to poor outcomes (Henry et al., 1990). Table 5:61 below shows the product-moment correlation coefficients, for each batch of

the British data set, between mean salience scores for each type of difficulty and weighted affiliation scores[51]. Successive rows in the table give the correlations for therapist as referent (surfaces 3, 1 and 2 respectively) and patient as referent (surfaces 1 and 2 respectively). Noteworthy associations (correlations above .14[52] and in the same direction across both batches) are shown in **bold** script, others are shown in *italic* script.

	MEAN SALIENCE SCORES					
	PARADIGMATIC		SITUATIONAL		TRANSIENT	
	B I	B II	B I	B II	B I	B II
Therapist Introject	**-.14**	**-.41**	*-.09*	*.10*	*.10*	*-.06*
Therapist active with patient	**-.14**	**-.23**	*.00*	*-.07*	*.14*	*.01*
Therapist reactive with patient	**-.14**	**-.39**	*-.14*	*.03*	*.18*	*-.10*
Patient active with therapist	*.07*	*-.10*	**-.24**	**-.30**	*.29*	*-.10*
Patient reactive with therapist	*.02*	*.04*	**-.21**	**-.23**	*.28*	*-.15*

Table 5:61 - Correlations between Affiliation Scores and Difficulty Types:

There are no noteworthy associations of affiliation scores with transient salience (in the last two columns). For paradigmatic salience (in the first two columns) all three associations with the therapist as referent are noteworthy, but neither of the associations with the patient as referent is. For situational salience (in the middle two columns) the reverse is true. Both patient-referenced associations are noteworthy, but none of the therapist-referenced ones is. Table 5:62 below shows the probabilities of the noteworthy associations per batch (columns 2 and 4) together with the t-values on which they are based (*columns 1 and 3)*, and the combined probabilities (column 6) together

---

[51] Untransformed scores were used, as inspection of histograms, box-plots and normal probability plots had shown the distributions of affiliation scores to approximate balanced normal distributions.

[52] Note that the previously set criterion of correlations above |.15| has been marginally lowered post-hoc in this instance, in order to capture the information contained in the data better.Table 5:62 confirms that this did not result in a threat to reliability.

with the z-values on which they are based (*column 5*). As it had been expected that difficulties would be associated with hostility (and therefore negatively correlated with affiliation scores), one-tailed tests have been used. As each association is one of five comparisons per difficulty type, which were then replicated, a criterion of p = .01 (p = .05 / 5) has been used to determine statistical significance. All the combined probabilities reach that criterion.

	PROBABILITIES (ONE-TAILED)					
	PARADIGMATIC SALIENCE					
df = 88	*t* *B I*	p B I	*t* *B II*	p B II	*z* ***combined***	**p combined**
Therapist Introject	*1.29*	.100	*4.21*	<.001	***3.85***	**<.0005**
Therapist active with patient	*1.31*	.100	*2.17*	.017	***2.43***	**.008**
Therapist reactive with patient	*1.31*	.100	*3.97*	<.001	***3.69***	**<.0005**
	SITUATIONAL SALIENCE					
df = 88	*t* *B1*	p B1	*t* *B2*	p B2	*z* ***combined***	**p combined**
Patient active with therapist	*2.28*	.013	*2.99*	.002	***3.69***	**<.0005**
Patient reactive with therapist	*2.03*	.023	*2.20*	.016	***2.96***	**.002**

Table 5:62 - Difficulty Type and Affiliation Score Correlations: Associated Probabilities

It is apparent from both tables, that paradigmatic difficulty salience is associated with hostility from therapists towards their patients and towards themselves, in the absence of perceived hostility from the patient. Conversely, situational difficulty salience is associated with hostility from the patient, both actively and reactively, in the absence of hostility perceived by therapists in themselves. Transient difficulty salience is not reliably related to hostility either in therapists or in patients.

An indication of the kind of transactions represented by the negative affiliation scores can be gleaned from correlations[53] with single cluster scores:

For **therapist introject**, the highest correlations of paradigmatic salience occur with **cluster 7** "Self-rejecting and destroying" (r = .18 / r = .35), followed by **cluster 6** "Self-indicting and oppressing" (r = .16 / r = .33);

for **therapist active** with patient, the highest correlation of paradigmatic salience occurs with **cluster 6** "Belittling and blaming" (r = .35 / r = .25);

for **therapist reactive** with patient, the highest correlations of paradigmatic salience occur with **cluster 7** "Protesting and recoiling" (r = .16 / r = .42), followed by **cluster 6** "Sulking and scurrying" (r = .18 / r = .28), and **cluster 5** "Deferring and submitting" (r = .15 / r = .26).

For **patient active** with therapist, the highest correlations of situational salience occur with **cluster 7** "Attacking and rejecting" (r = .17 / r = .34), followed by **cluster 6** "Belittling and blaming" (r = .17 / r = .30), and **cluster 5** "Watching and controlling" (r = .14 / r = .14);

for **patient reactive** with therapist, the highest correlation of situational salience occurs with **cluster 7** "Protesting and recoiling" (r = .12 / r = .20).

It appears that therapists' transactions during paradigmatic difficulties are even more characterised by hostile enmeshment (interdependence) - that is by an aggressive wish to control - than by simple disaffiliation. This would be consistent with a view that therapists, in this type of difficulty, strive for the mastery of an internal conflict both in themselves (introject) and in their patients; and that their antagonism is generated by this internal conflict rather than by the interactions with their patients.

---

[53] Correlations are only quoted if they are positive and above .1 in both batches

Some further corroboration is provided by the correlations of weighted autonomy scores with difficulty types.[54] The most telling results here are the associations with the '**therapist reactive** with patient' surface. For transient salience they are inconclusive (r = -.12 / r = .00), for **paradigmatic** salience they are **negative** (r = -.20; df = 88; t = 1.88; p = .06 / r = -.39; df = 88; t = 3.91; p < .0005 | combined probability z = 4.01 ; p < .0005), but for **situational** salience they are **positive** (r = .27; df = 88; t = 2.63; p = .01 / r = .22; df = 88; t = 2.10; p = .04 | combined probability z = 3.31; p < .0005). Therapists experiencing paradigmatic difficulties have a tendency to deny autonomy ("Watching and controlling") despite the absence of patient hostility. In sharp contrast, therapists experiencing situational difficulties can promote autonomy ("Freeing and forgetting"), even in the face of hostile attacks from their patients, because, one might speculate, their reactions are not constrained by the need to try and manage past conflicts in a current situation.

Finally, in this section, associations of difficulty types with conflict coefficients are of interest. These are also bi-modally distributed (like attack and control coefficients) but convey quite different information dependent on whether they are positive or negative. They can therefore be analysed separately as variables with unimodal (and usually approximately normal) distributions. Positive conflict coefficients denote interdependence conflicts, whereas negative conflict coefficients, which occur much less frequently, are indicative of attachment (love/hate) conflicts. Alpher (1996) suggests that introject conflict coefficients are a direct measure of intrapsychic conflict and demonstrates their discriminative power in relation to patients' diagnostic categories. For the current enquiry, difficulty-specific introject conflict coefficients can be

---

[54] The complete table is provided in appendix 16

understood as a measure of therapists' intrapsychic conflict in the course of a difficulty. As such, they should be associated with paradigmatic difficulty salience, but not with transient difficulty salience. Given the generally high levels of affiliation displayed by therapists, it is likely that negative conflict coefficients will be of greater importance. Two hypotheses can be derived from this line of argument:

1. Negative conflict coefficients for difficulty-specific therapist introjects will be negatively correlated with paradigmatic mean salience scores, but not negatively correlated with transient mean salience scores.
2. Positive conflict coefficients for difficulty-specific therapist introjects will be positively correlated with paradigmatic mean salience scores, but not positively correlated with transient mean salience scores.

Table 5:63 below shows the correlations of conflict coefficients with mean salience scores for the three difficulty types across the two batches of data. As conflict coefficients are themselves correlation coefficients (between an observed array of cluster scores and an ideal, theoretical distribution), Fisher z transformations have been applied. Noteworthy associations are shown in **bold** typeface, others in *italic*.

CONFLICT COEFFICIENTS: (Fisher z transformed)	MEAN SALIENCE SCORES: PARADIGMATIC B I	PARADIGMATIC B II	SITUATIONAL B I	SITUATIONAL B II	TRANSIENT B I	TRANSIENT B II
Attachment Conflict (negative coefficients) N = 18 / N = 23	**-.21**	**-.33**	**.19**	**.31**	*.22*	*.07*
Interdependence Conflict (pos. Coeff.) N = 72 / N = 67	*.03*	*-.08*	*.05*	*.02*	*-.03*	*.24*

Table 5:63 - Therapist Introject Conflict Coefficients: Correlations with Difficulty Types

It is apparent that there are no noteworthy associations between positive conflict coefficients and mean salience scores. The second hypothesis is therefore unsupported.

There are, however, in line with predictions, noteworthy negative associations between negative conflict coefficients and paradigmatic mean salience scores, in the absence of such noteworthy negative associations with transient mean salience scores. There is also a noteworthy positive association between negative Conflict Coefficients and situational mean salience scores, which had not been predicted. As the number of occurrences of attachment conflicts is small (N = 18 for Batch1 / N = 23 for Batch2), amounting to only 20% - 25% of all cases, the probabilities associated with the noteworthy associations fail to reach statistical significance (only marginally for paradigmatic salience, but - because of the use of a two-tailed test - by a wider margin for situational salience)[55]. The first hypothesis is therefore also not supported. Nevertheless, the results are consistent with the view that paradigmatic difficulties are associated with some intrapsychic conflict (an ambivalence about friendly or hostile self-attachment), while transient difficulties are not; and that situational difficulties, being by definition located

---

[55] Attachment conflict * paradigmatic salience: r = -.21; df = 16; t = .84; p (one-tailed) = .21 / r = -.33; df = 21; t = 1.58; p (one-tailed) = .065 / combined probability: z = 1.61; **p (one-tailed) = .054**. Attachment conflict * situational salience: r = .19; df = 16; t = .76; p (two-tailed) = .46 / r = .31; df = 21; t = 1.49; p (two-tailed) = .15 / combined probability: z = 1.5; **p (two-tailed) = .133.**

outside the therapist, are characterised by the absence of such intrapsychic conflict. This interpretation is limited, though, by the fact that such ambivalences are occurring relatively infrequently.

5.3.5.2.2 Associations with Measure of Distance

Respondents to TDQ2 were not asked to speculate about the introjects of their patients during the course of the difficulties they described. It is therefore not possible to compare therapist and patient introjects in order to replicate the findings obtained from IIP data concerning the greater similarity between therapists and those patients with whom they experience more paradigmatic difficulties (see 5.3.4.2 above). It is, however, possible to operationalise directly the hypothesis derived from research question 3 - that paradigmatic difficulties will be associated with therapists' enduring characteristics while transient difficulties will not - by comparing therapists' introject ratings. If a difficulty had its origins in the therapist's personal characteristics, one would expect the difficulty-specific introject to be related to the general introjects (which should be enduring). In particular, one would expect the difficulty-specific introject to be similar to the general introject 'at worst', because both relate to problem states. Conversely, one would expect the difficulty-specific introject to be dissimilar to the general introject 'at best', because the latter should reflect a state relatively free of conflicts. Quintana & Meara (1990) suggest the Euclidean distance between combined affiliation and autonomy scores as a suitable measure of dissimilarity between surfaces. Each therapist introject is characterised by a specific point on the circular surface whose co-ordinates are defined by its affiliation score (on the horizontal axis) and its autonomy score (on the vertical axis). Distance between two introjects I and II is then defined as the square root of the sum of the squared differences between affiliation scores I and II and

autonomy scores I and II[56]. The relationship of this measure of dissimilarity to paradigmatic mean salience scores can then be investigated and hypotheses formulated as follows:

1. Distances between difficulty-specific therapist introjects and therapist introjects 'at best' will be positively correlated with paradigmatic mean salience scores, but not positively correlated with transient mean salience scores.

2. Distances between difficulty-specific therapist introjects and therapist introjects 'at worst' will be negatively correlated with paradigmatic mean salience scores, but not negatively correlated with transient mean salience scores.

Table 5:64 below shows the correlations of distances between difficulty-specific and general introjects with mean salience scores for the three difficulty types across the two batches of data. Noteworthy associations are shown in **bold** typeface, others in *italic*.

EUCLIDEAN DISTANCES:	**MEAN SALIENCE SCORES:**					
	PARADIGMATIC		**SITUATIONAL**		**TRANSIENT**	
	B I	B II	B I	B II	B I	B II
Difficulty Introject - Introject at best	**.19**	**.22**	*-.01*	*.09*	*-.05*	*.07*
Difficulty Introject - Introject at worst	**-.18**	**-.26**	*-.02*	*.03*	*.26*	*.01*

Table 5:64 - Distances between Introjects: Correlations with Difficulty Types

It is readily apparent that the only stable and substantive correlations occur with paradigmatic salience scores. Both are in the predicted direction. Table 5:65 below shows the single batch (B I and B II) and combined probabilities for both introject comparisons together with the associated *t* or z values (in *italics*).

---

[56] $D = \sqrt{(af_I - af_{II})^2 + (au_I - au_{II})^2}$; af = affiliation scores, au = autonomy scores

df = 88	PARADIGMATIC SALIENCE PROBABILITIES (ONE -TAILED)					
	t *B I*	p B I	*t* *B II*	p B II	*z* ***combined***	**p combined**
Difficulty Introject - Introject at best	*1.84*	.035	*2.13*	.018	***2.78***	**.001**
Difficulty Introject - Introject at worst	*1.71*	.045	*2.57*	.006	***2.99***	**.001**

Table 5:65 - Correlations between Introject Distances and Difficulty Type: Probabilities

The associations, though small, are highly reliable. Both of the above hypotheses are therefore supported and the results provide further evidence for the connection between paradigmatic type difficulties and enduring personal characteristics of the therapist who experiences them, as well as for the absence of such a connection in transient and situational type difficulties.

5.3.5.3 Summary of Results and Discussion

The analyses in the preceding section explored therapist / patient interpersonal transactions during difficulty situations and therapists' resultant specific introjects compared with their general introjects 'at best' and 'at worst'. They allowed for testing of two different operationalisations of the hypothesis that paradigmatic difficulties are linked to therapists internal conflicts.

Therapist introjects were found to differ widely between best and worst states. Therapists in both language-samples locate their best states in the right, self-loving half of the surface 1 circumplex (with a peak in the 'self-accepting and exploring' octant), while their worst states are characterised by hostile self-control. Difficulty-specific introjects lie in the main between these extremes, but are generally closer to the best state. This similarity is pronounced for not-so-difficult, but attenuated for difficult patients. In their private lives, on this evidence, therapists 'at best' treat themselves as

well as comparable groups, while 'at worst', they are probably more given to being hostile and coercive towards themselves than normal comparison groups (though not as much as clinical samples). When in role, therapists maintain a basically friendly attitude towards themselves, even when they feel challenged professionally. This stance is harder to maintain, however, with patients regarded as 'difficult', or, alternatively (and probably more likely), patients are experienced as difficult if they succeed in challenging their therapists' positive self-image. On the other hand, professional challenges markedly cause therapists to monitor and restrain themselves, in sharp contrast to their usual attitude when not in role and in their best state. However, as no baseline measures were taken for therapists when in role but not in difficulty, it remains unclear how much this self-restrictive stance would generally be characteristic of their working selves. One might speculate that there would be a wide range, probably associated with theoretical orientation, which then narrows considerably in the face of difficulties in practice.

As far as therapist / patient transactions are concerned, therapists see themselves as acting and reacting in a friendly and encouraging manner. They experience their difficult patients as hostile, both in approach and reactions, and their not-so-difficult patients as friendly but autonomous in approach, and assertive/autonomous in reaction. These features are broadly similar for English- and German speaking samples. Associations between transactions and patient categories allow these features to be organised into the following (speculative) sequence: Therapists (on the strength of their perceptions) approach their patients generally in a friendly and encouraging manner during difficulty situations. Their difficult patients treat them with hostility, their not-so-difficult patients with affirmation and moves towards separation. In response, therapists

maintain a consistent friendly but watchful attitude towards both kinds of approach, which results in more affiliative reactions from their patients, leading to a convergence between those experienced as difficult and those as not-so-difficult.

This picture is in sharp contrast to that emerging from direct associations between weighted affiliation scores and difficulty types[57]. Paradigmatic salience is consistently and reliably associated with therapist hostility on all three surfaces in the absence of perceived hostility from the patient. Even though therapists generally manage to treat their hostile, difficult patients in a friendly manner, they act and react with hostile control towards patients with whom they experience paradigmatic difficulties (even though these show no particular hostility). As therapists also attack themselves in these situations, a plausible explanation for these findings would be that while experiencing paradigmatic difficulties they attempt to master an internal conflict in themselves and project this struggle on to their patients, whose interpersonal transactions would not otherwise warrant such a reaction. Situational difficulties are associated with the opposite state of affairs. In these situations, therapists maintain a friendly attitude towards themselves and toward their patients whom they experience as generally hostile. As they are not compelled to attempt controlling their own internal conflicts in their patients, they can maintain a stance which encourages differentiation and autonomy. Although conjectural, the preceding interpretations are consistent both with the empirical data and with the theoretical concepts of paradigmatic and situational difficulties.

As SASB provides a direct measure of conflict (tapping high endorsements on opposing poles of the affiliation or interdependence axes), the general hypothesis that such

---

[57] These results are confined to the British sample as salience ratings for the German-language sample were not available.

conflicts in therapists would be associated with paradigmatic difficulties could be tested directly. No such association was found for the (relatively frequently occurring) interdependence conflicts in therapists' introjects. Attachment (self-love vs. self-hate) conflicts, on the other hand, showed a consistent modest association with paradigmatic salience. Owing to the low incidence of such conflicts, these correlations narrowly miss statistical significance at the .05 level. Although the hypothesis is formally unsupported, the results contribute to accumulating evidence for the link between therapists' internal conflicts and paradigmatic difficulties.

Another test of the hypothesis was derived from operationalising internal conflicts in terms of distances between difficulty-specific and baseline introjects. Positive associations of paradigmatic salience with distances to baseline introject 'at best' and negative associations with those to introject 'at worst', were consistent and reliable in line with predictions. On these measures the general hypothesis was supported.

In summary, results from SASB INTREX repeat in difficulty-specific terms the findings provided by IIP-51 in general terms. Enduring similarities between therapists and patients translate into situational conflict, and both are associated with paradigmatic difficulty salience.

5.3.6 CCRT Grid Data

The grid using CCRT standard categories as elements and constructs was designed to provide direct evidence of therapist / patient similarity (see 5.2.1.1.7.6) above. In the event, this part of TDQ2 failed entirely. Only 75% of British and 70% of English-language respondents completed the relevant section of the questionnaire. Of those who did supply data, the majority reported in the feedback section, that they had experienced difficulties with the grid. Many spoke of their confusion and it was evident from some comments that respondents had misunderstood the explanations. Not entirely surprisingly, British clinical psychologists seem to have coped best. One aspect frequently attracting negative comments was the composite nature of some of the standard categories (for instance "someone wishing to oppose, control, and hurt others" or "self-controlled and self-confident") leading some respondents to attempt differential ratings of the category components. The only area in which this measure evidently succeeded was in evoking negative and often exasperated feelings in respondents. Fortunately, in view of its complexity the grid had been placed at the end of TDQ2; otherwise it might have strained the alliance with respondents to breaking point before they had completed other parts of the questionnaire. In retrospect, it is clear that the measure had not been piloted carefully enough. Asking someone not particularly invested in research and unfamiliar with repertory grid methodology to complete the measure before including it in TDQ2, might have revealed its shortcomings early on and saved both researcher and respondents a great deal of time and trouble[58].

---

[58] The covering letter to be sent out with a summary of the findings to those respondents who have requested it, will take up the feedback received and apologise for the inclusion of this measure.

5.3.7 Therapist Self-Evaluations

Although items relevant to this section follow on immediately from the FID scales in TDQ2, results are considered here last, because they are better understood in the light of already established convergent validity of the difficulty types. Analyses in this section draw on the two items enquiring about the specificity/universality of the difficulty experience (one - item 33 - within therapists across situations, the other - item 32 - across therapists for the same situation) and on seven items (item 30 a) - c) and item 31 a) - d)) assessing deficits in skills, knowledge and experience.[59] These latter items provide the means of testing the second general hypothesis (see 5.1.2 above) which is central to validating the transience concept.

5.3.7.1 Competency Deficits

Although deficits are grouped under two different headings, the question arises whether the items do in fact form a single scale. Means and standard deviations (in brackets) for all deficit items, differentiated by patient categories and data sets, are shown in Table 5:66 below. The small Ns for the German data set are the result of the relevant page having accidentally been omitted from some of the German questionnaires, which disproportionately affected accounts of difficulties with not-so-difficult patients. Means are calculated from ratings on a six-point (0 - 5) scale with '0' denoting a quality which would not have been useful or a resource which would not have been helpful and therefore a low deficit rating.

---

[59] The two open-ended items 30 d) and 31 e) were omitted as they had been positively endorsed by less than a third of all respondents.

	English Data Set N = 100		German Data Set N = 24 / *19*	
	DP	***NSD***	**DP**	***NSD***
Getting formal supervision or consultation	**4.1** (1.4)	***3.7*** *(1.6)*	**3.4** (1.4)	***3.2*** *(1.5)*
A wider range of practical experience	**3.0** (1.7)	***2.7*** *(1.6)*	**3.3** (1.6)	***2.8*** *(1.8)*
Reading relevant literature	**2.6** (1.5)	***2.1*** *(1.5)*	**2.0** (1.5)	***1.6*** *(1.7)*
Being familiar with wider range of clients	**2.6** (1.7)	***2.1*** *(1.7)*	**3.0** (1.6)	***2.5*** *(1.7)*
Attending workshops or seminars	**2.3** (1.6)	***2.1*** *(1.5)*	**1.3** (1.4)	***2.0*** *(1.7)*
A broader repertoire of technical skills	**2.2** (1.6)	***1.8*** *(1.6)*	**1.9** (1.6)	***1.9*** *(1.7)*
More extensive theoretical knowledge	**2.0** (1.7)	***1.8*** *(1.6)*	**1.7** (1.5)	***1.9*** *(1.7)*

Table 5:66 - Competency Deficit Items: Means and Standard Deviations

Formal supervision or consultation was most frequently identified as the missing resource which would have been helpful in equipping therapists to deal with the difficult situations they were describing. Rank orderings of items are fairly stable across samples, with theoretical knowledge being the least endorsed deficit. Respondents also took a generally sceptical view of the value of having broader technical skills. This result may have been influenced by the preponderance of experienced therapists in the sample, who may have felt that their technical repertoires were already broad enough. German therapists were somewhat more inclined to find familiarity with a wider range of clients useful, and less positive about the value of workshops or seminars to help with difficult patients. However, the small sample size militates against interpreting these results. [60]

---

[60] A subsidiary, but nevertheless interesting, question concerns possible relationships between theoretical frames and perceptions of particular deficits. Examining the intercorrelations between five theoretical influence ratings (psychodynamic, behavioural, cognitive, humanistic and systemic) and the seven deficit items, two associations emerged as noteworthy (correlations

Principal component analyses (with subsequent Varimax rotations) extracted either one or two factors with eigenvalues > 1, depending on the sample. Furthermore, the assignment of items to factors in the two-factor solutions was not stable across samples. It was therefore decided to treat all seven items as a single scale and to calculate scale alphas. The item "getting formal supervision or consultation" was found to depress scale alphas across all samples and was therefore excluded. The resulting six-item "Deficit Scale" was found to be highly internally consistent and therefore suitable for use in further analyses. Descriptive statistics for the scale, together with Cronbach's alphas, across samples are displayed in Table 5:67 below.

	English Data Set N = 100		**German Data Set N = 24 / *19***	
	DP	**NSD**	**DP**	**NSD**
Mean	**2.41**	**2.12**	**2.19**	**2.11**
SD	1.19	1.18	1.13	1.36
Median	*2.67*	*2.17*	*2.17*	*2.00*
Range	*0 - 5*	*0 - 5*	*0 - 4.8*	*0 - 4.8*
Scale α	**.82**	**.84**	**.84**	**.88**

Table 5:67 - Deficit Scale: Descriptive Statistics and Reliabilities

Inspection of histograms and normal probability plots showed deficit scale scores to be approximately normally distributed around the means shown in the first row of the table. Medians, displayed in row 3, are very close to the means, confirming that sample distributions are not unduly skewed. Standard deviations and ranges, shown in the

---

above .15 in the same direction across both batches of data): The more that therapists are influenced by psychodynamic theory, the more they perceive a lack of having read relevant literature (r = .23; t = 2.4; df = 98; p = .02 / r = .21; t = 2.1; df = 99; p=.04 / combined prob. z = 3.15, p < .001); and the more therapists are influenced by behavioural theory, the less they perceive a lack of supervision and consultation (r = -.21; t = -2.1; df = 97; p = .04 / r = -.19; t = -2.0; df = 99; p = .05 / combined prob. z = 2.83; p = .003). Even allowing for the fact that these were the only two noteworthy associations out of a total of 35 that were investigated, the associated combined probabilities ar low enough to raise questions about the coping resources promoted in either framework.

second and fourth row, indicate adequate variability. Finally, the alphas in the last row are satisfactory and establish the utility of the scale in both English and German versions.

Means are very similar in three of the samples, with average endorsements just above the second of the five scale intervals. British respondents, however, feel their deficits more keenly when dealing with difficult patients, resulting in a mean endorsement just below the midpoint of the scale. The difference between the means in respect of difficult and not-so-difficult patients is also statistically significant for the English sample ($t = 3.0$; $df = 98$; $p = 0.004$).

5.3.7.2 Universality and Specificity Items

Descriptive statistics for the single items tapping universality and specificity are shown in Table 5:68 below. Inspection of normal probability plots and histograms showed the item scores to be approximately normally distributed around the means shown in the first row of the table, with the exception of the specificity item in the German 'difficult patient' sample which shows a bi-modal distribution within a narrow range. Figures for this sample are given in square brackets as they do not lend themselves to statistical comparisons. Why this sample should behave so differently from the others is unclear; however, the small N makes the German sample generally a less reliable source of information for these items.

	UNIVERSALITY (Item 32) (How **difficult** would another therapist, of a theoretical orientation and a level of experience similar to yours, have found the situation?)				SPECIFICITY (Item 33) (How **often** have you experienced a difficulty of this kind in a situation outside therapy (that is in a family, social or work relationship)?)			
	English Data Set N = 100		German Data Set N = 24 / 19		English Data Set N = 100		German Data Set N = 24 / 19	
	DP	NSD	DP	NSD	DP	NSD	DP	NSD
Mean	**3.87**	**3.17**	**4.04**	**3.11**	**1.67**	**2.06**	**[.88]**	**1.95**
SD	.92	1.02	.81	1.10	1.26	1.27	[.95]	1.55
Median	*4.0*	*3.0*	*4.0*	*3.0*	*1.0*	*2.0*	*[1.0]*	*2.0*
Range	*2 - 5*	*0 - 5*	*2 - 5*	*1 - 5*	*0 - 4*	*0 - 5*	*[0 - 3]*	*0 - 5*

Table 5:68 - Universality and Specificity Items: Descriptive Statistics

Excepting the sample mentioned above, means are quite similar across language sets. Medians, which are less vulnerable to extreme values, are identical. Universality means for difficult patients are somewhat elevated, which is not surprising given that situational difficulty salience, which should be predicted by higher universality scores, is positively associated with this patient category (see 5.3.2.1).

5.3.7.3 Intercorrelations between Self-Evaluation Measures

The next question that arises concerns the intercorrelations between deficit means and universality/specificity scores, which are shown in Table 5:69 below. Figures in normal script, below the diagonal relate to 'difficult patient' samples. Figures in *italic script*, above the diagonal, relate to 'not-so-difficult patient' samples. **Bold** figures, on the upper left of each cell relate to English-language samples, figures to the lower right of each cell relate to German-language samples. Figures in square brackets relate to the bi-modally distributed sample and are given only for the sake of completeness.

	Deficit	Universality	Specificity
Deficit	1	***.22*** *.23*	***-.01*** *.17*
Universality	**.14** -.11	1	***.01*** *.10*
Specificity	**.03** [.55]	**-.17** [-.11]	1

Table 5:69 - Intercorrelations between Deficit, Universality and Specificity Items

The safest conclusion from these results appears to be that the three variables are substantially unrelated to each other, although there is a weak positive association between Deficit and Universality in three out of the four samples. Consequently, they might be suitable predictor variables in a regression model, as they are unlikely to raise concerns about multi-collinearity.

5.3.7.4 Associations with Difficulty Types

The important question in relation to self-evaluation of difficulties by respondents, however, concerns the relationships of the three variables with the difficulty types. The deficit scale is the only measure in TDQ2 which assesses therapists' skills. The second general hypothesis is therefore operationalised as follows:

- Deficit scale mean scores relating to a difficulty will be positively correlated with mean transient salience scores but not with paradigmatic mean salience scores for that difficulty.

As outlined in 5.2.1.1.4 above, if therapists were able to classify their difficulties accurately, one would further expect

a) transient salience to be positively associated with Universality but negatively associated with Specificity;

b) situational salience to be positively associated with Universality but negatively associated with Deficit and Specificity; and

c) paradigmatic salience to be positively associated with Specificity but negatively associated with Deficit and Universality.

Correlation coefficients for the relationships across both batches of data between mean salience scores for types of difficulty on the one hand, and Deficit means, Universality and Specificity scores on the other, are shown in Table 5:70 below. Noteworthy correlations are shown in **bold** type[61].

	MEAN SALIENCE SCORES					
	PARADIGMATIC		**SITUATIONAL**		**TRANSIENT**	
	Batch I	*Batch II*	Batch I	*Batch II*	Batch I	*Batch II*
Deficit	-.06	*-.10*	-.09	*.04*	**.33**	***.36***
Universality	-.05	*-.04*	**.34**	***.25***	-.16	*.05*
Specificity	**.25**	***.26***	**-.18**	***-.14***	-.19	*.01*

Table 5:70 - Therapists' Self-Evaluations: Associations with Difficulty Types

It appears that each of the difficulty types is most strongly associated with one 'marker variable', though situational salience shows a consistent but weaker association with a second variable. The largest correlation (mean r = .345) is that between transient salience and Deficit, which is also highly reliable ($t = 3.38$; $df = 93$; $p < .001$ / $t = 3.76$; $df = 94$; $p < .001$ / $z = 5.00$; **p(comb) < .0001** (one-tailed test)). There is a consistent, but very weak and unreliable negative association between paradigmatic salience and Deficit. The hypothesis is therefore supported in both parts.

[61] As before, the criterion has been marginally lowered post-hoc to correlations above |.14|, in order to include what seemed likely to be a reliable association.

The other noteworthy associations are more modest, but in the directions predicted and reliable as follows:

Paradigmatic & Specificity: $t = 2.51$; $df = 93$; $p = .007$ / $t = 2.64$; $df = 94$; $p = .005$

$z = 3.60$; **p(comb) < .0005** (one-tailed test).

Situational & Universality: $t = 3.42$; $df = 90$; $p < .001$ / $t = 2.50$; $df = 90$; $p = .007$

$z = 4.14$; **p(comb) < .0005** (one-tailed test).

Situational & Specificity: $t = 1.72$; $df = 93$; $p < .040$ / $t = 1.39$; $df = 94$; $p = .083$

$z = 2.18$; **p(comb) = .015** (one-tailed test)[62].

Regression models can be fitted reliably across the two batches for two of the difficulty types as follows. Item combinations are in line with theoretical expectations, but resultant multiple correlations are only marginally greater than the associations with single variables shown above:

Paradigmatic Salience = Constant + Specificity - Deficit

multiple $r = .26$; $F = 3.28$; $p = .04$ / multiple $r = .28$; $F = 3.89$; $p = .02$

Situational Salience = Constant + Universality - Specificity

multiple $r = .36$; $F = 6.79$; $p = .002$ / multiple $r = .27$; $F = 3.60$; $p = .03$

Inspection of probability plots showed residuals to be normally distributed, indicating the absence of other significant predictor variables.

The item "getting formal supervision or consultation", which had been excluded from the deficit scale does not show any noteworthy associations with difficulty types[63].

---

[62] As three comparisons were involved, the criterion for statistical significance was set at $p < .017$ (.05 / 3).

[63] It does, however, show a moderate negative association with practice length, which is quite independent from the positive correlation with behavioural orientation mentioned above. Both these variables can therefore be entered into a regression model, which is reliable across batches: Supervision = Constant + Practice Length + Influence of Behavioural Orientation

multiple $r = .39$; $F = 8.33$; $p < .0005$ / multiple $r = .33$; $F = 6.03$; $p < .004$

5.3.7.5 Summary of Results and Discussion

The analyses in the preceding section used three instruments - a six-item competency deficit scale and two single-item measures, one tapping the universality of an experience of difficulty across different therapists, the other its specificity for a particular therapist (expressed as its pervasiveness across different situations). The Deficit scale combines items based on absence of resources with items tapping insufficiencies in skills, knowledge, and experience. It shows good variability and its internal consistency has been demonstrated across patient categories and language-communities. Both the Universality and Specificity items discriminate well, are independent of each other, and are largely independent of the Deficit scale. Consequently, there is reason to have confidence in the three measures as serviceable dependent variables.

The hypothesis that transient difficulties would be associated with therapists' deficits in skills, knowledge, and experience was borne out on the basis of these self-report measures. More surprising was how well the single Specificity item performed. It serves, as theoretically expected, as a marker variable for paradigmatic difficulties, but also differentiates between these and situational difficulties. The single Universality item showed the expected positive association with situational, but not with transient difficulties. This raises the question, whether two items would have represented the relevant difficulty type definitions better - one in the present form, and another asking, how difficult other therapists, regardless of their orientation or experience, would have found the situation described.

Experienced behavioural therapists, it appears, are even less likely than their younger colleagues to think of formal supervision or consultation as a resource for dealing with their difficulties.

Generally, it would seem useful to expand the Universality and Specificity items into short scales, if one wanted to improve the self-evaluation section of the questionnaire or was interested in constructing a brief self-diagnostic instrument for the use of clinicians. Further reliability might be gained by making use of the decision-making rules contained in the rating instructions (see 4.2.4.2) and adding questions about the location of the difficulty either inside or outside the therapist. It would also be informative to supply therapists with the difficulty type definitions, to investigate how much their own evaluations are congruent with those of external judges.

As respondents to TDQ2 were not asked questions tapping difficulty type directly, evidence of their awareness of the kind of difficulty they were experiencing is only indirect. Even allowing for attenuation by measurement error, associations between difficulty types and marker variables are modest. Inspection of scatterplots did not reveal clusters of high or low associations, indicating that therapists are approximately normally distributed regarding the accuracy of their self-evaluations. The most likely interpretation of these results would be that therapists were vaguely cognisant of the levels of competence deficits, universality and specificity entailed in their accounts; that these features were either marginally conscious in their minds, or else on the threshold of awareness (pre-conscious). One might therefore conclude that a supervisory strategy focusing on the criteria distinguishing difficulty types might bring these fully into awareness quite readily - at least with therapists as open to self- examination as the respondents to TDQ2 evidently were.

5.4 Discussion

The study based on the survey with TDQ2 was involved, complex, and costly for respondents and researcher alike. This lends poignancy to the customary reflections on what it accomplished, what it failed to achieve, what could have been done differently, and what might be done in the future. Deliberations fall into three domains: a) measures and design, b) research questions and hypotheses, and c) incidental discoveries about difficult patients.

First among the new measures constructed for TDQ2 were the shortened Familiarity and Impact scales. Although specifically designed for difficulties in therapy, they seem likely useful measures of experiential aspects of any challenging event, evincing promising psychometric properties and demonstrated cross-cultural validity. The Subjective Duration scale did not appear to make a useful contribution in its current form, but might be improved by dropping items confounded with impact, and adding to those which tap the brief/enduring dimension. Ipsatisation of scores clearly sharpens separation between the scales and would be useful in studies aiming to assess differential effects of challenging events.

The second set of newly designed measures concerned therapists' self-reported competency-deficits, universality, and specificity in relation to a difficulty. The Deficit scale, if shortened by one item, is internally consistent in both German and English versions. If one thinks of difficult situations as 'critical incidents', testing the limits of a therapist's skills, knowledge, and experience, scores aggregated from several such situations might well provide a general index of therapist competence level, especially when controlling for paradigmatic salience. The single-item Universality and Specificity measures have emerged as practicable questions to ask in supervison or private

reflection, in order to clarify locus and type of a difficulty. Further work might usefully develop them into short scales.

The last of the newly designed instruments, the CCRT grid, did not succeed for the reasons outlined in 5.3.6 above. Given the problems respondents experienced in understanding instructions and completing the measure, there seems little hope of improving it. Had this section of TDQ2 been piloted more carefully, its flaws might have been noticed earlier, leading to a shorter, more user-friendly questionnaire. While the idea of assessing core conflicts in relation to paradigmatic difficulties still appeals, the detailed interpersonal formulations of the SASB model appear to hold greater promise than the more global and less well differentiated CCRT standard categories[64].

Of the established measures, SASB INTREX is specifically designed to be used flexibly and adapted to particular research questions. Its utility in the format used for the current study was therefore expected. The IIP-51, on the other hand, is a departure from previously published forms of the inventory. Although there seems to be an ever growing list of IIP versions, the richness of information captured by this instrument might make it worthwhile to add it as yet another, provided an examination of the lack of overlap between its circumplex and factor-based components with a clinical sample would replicate the results from this study.

The prompts with which difficulty accounts were elicited proved successful in securing extended narratives and were an improvement on those in TDQ1. Asking for difficulties with 'difficult' and 'not so difficult' patients led to interesting unplanned observations which are discussed below. The rating instructions were confirmed as being quite

---

[64] One of the authors of the previous attempt to design a CCRT self-report format has now abandoned the idea of tapping unconscious wishes or intentions, and is experimenting instead with lists of reactions based on SASB fomulations (R. W. Dahlbender, personal communication, June 1998).

adequate, although they could be further improved by introducing tighter decision-making rules. Reliability for the transient difficulty type is likely to have suffered from withholding of information about length of therapists' practical experience. Now that the link of transient salience with earlier career stages has been established, further studies should aim to provide as much contextual information as is feasible. Although inter-rater reliabilities were generally satisfactory (and quite substantial for paradigmatic and situational salience), they were achieved with a group of five raters. Improvements in the rating instructions and enhanced contextual information might be expected to decrease the number of raters necessary for reliable judgements and bring estimated single-rater reliabilities up to acceptable levels.

Although TDQ2 was very lengthy, substantial numbers of respondents did complete it and some even commented positively on the experience (despite the CCRT grid)[65]. The opt-in procedure also attracted favourable comments and it might be assumed that it contributed to the research alliance with respondents. However, its impact is as hard to assess as that of other features aimed at forming and fostering such an alliance. It is noticeable though, that the Swiss data collection among CRN members, whose previous engagement had not been maintained by promised newsletters, resulted in complete lack of response.

Successes in reaching and engaging potential participants should not obscure the fact that overall response rates were very low, especially when expressed as a percentage of target groups rather than distributed questionnaires. This raises questions about the kind of self-selection processes that might have operated, limiting the generalisability of any

---

[65] One German respondent commented (freely translated): "I have always finished what I started in my life, so I have finished this questionnaire, but this is the last one I am doing for the next five years!". This resolve gladdens the heart of the therapist but saddens that of the researcher.

findings. The willingness to write detailed narratives of therapy events (not far short of 1000 words in some instances), speaks of a disposition to introspection and a capacity for reflection in respondents. Their tenacity in completing the rest of the questionnaire demonstrates tolerance to frustration and maybe an over-affiliative stance, after all respondents' self-descriptions in terms of the IIP marked them out as co-operative to the point of gullibility. However, the proximity of these descriptions to that of a helping-professional prototype gives some confidence that respondents are not too unusual when compared with their peers. Maturity as a group, both in terms of age and practical experience, and an investment in psychodynamic and humanistic theoretical frameworks, emerge from the sample characteristics. Mean impact ratings above the scale mean reveal respondents as being affected by the difficulties they report, maybe echoing the finding from TDQ1, showing raised overall levels of experiences of difficulty among this group. Overall, the picture of a typical respondent that emerges, is that of a therapist who I would not mind having in supervision with me, and who I would not mind taking my own cases to, as a peer. Given that the impetus for this whole line of enquiry arose out of supervisory experiences, the limitations imposed by the self-selected nature of the sample may not be too critical.

The potential of the design to generate cross-cultural understandings is limited by the lack of difficulty type ratings for the German accounts. The low number of respondents militated against expending the effort of recruiting and training German-language raters, or of validating translations of German narratives. However, the findings relating to difficulty type would be strengthened considerably if they could be shown to be replicated (as findings derived from the various other measures frequently have been).

Further work might usefully concentrate on securing a larger pool of German-language difficulty accounts to make a rating effort worthwhile.

As a result of this third study, the three difficulty types and their definitions have been validated. Transient and paradigmatic difficulties have emerged as independent aspects of the therapist's experience and their respective links to competency deficits and enduring therapist characteristics confirmed. Situational difficulties have emerged as a valid third category, linked to some degree to patients experienced as difficult. Evidence for the validity of the categories is cumulative and convergent. Of the measures employed, only the CCRT grid failed to contribute to this result because it proved to be unusable. Within the FID measures, the duration indices failed to relate to difficulty type, while impact and familiarity are associated with paradigmatic and transient salience respectively.

It might be argued, that despite their reliability, many of the observed associations which were in line with predictions are quite modest. Effect size is arguably more important than statistical significance (Cohen, 1990). However, conducting a survey endeavouring to capture clinical reality is different from manipulating experimental conditions in a tightly controlled design, and effect sizes are likely to be attenuated by 'noisy' measures. The strategy of paying attention to 'noteworthy' associations before testing for statistical significance, was an attempt to give proper consideration to effect sizes, though the criterion adopted was low. On the other hand, very small effect sizes can be informative when applied to epidemiological questions (Rosenthal, 1990), despite the small amount of variance explained. Each reliable individual association with difficulty types observed in this study is small enough to render it ineffectual for predictions in individual cases, however, cumulatively they present a strong argument

for the validity of the theoretical concepts which generated the definitions for transient, situational, and paradigmatic difficulties. Further work with the current data set might usefully investigate the relationships of salience ratings and other measures within individual respondents.

The same strategy would also shed further light on the ways in which general perceptions of difficult and not-so-difficult patients are translated into situation-specific ones. The suggestion, derived from group means, that not-so-difficult patients are behaving out of character, by being assertive and challenging, when causing difficulties for their therapists, seems worthy of further investigation. Generally, perceptions of difficult and not-so-difficult patients are stable, distinct and worthy of reporting as, albeit incidental, findings in their own right.

6. Conclusion

At the end of this long process, what do we know better now than we did at the beginning, and why is it worth knowing?

Therapists' difficulties had previously been investigated in terms of their *content* and their *prevalence*, leading to findings about their *location* (in therapists, in patients, or in external circumstances) and their *relationship to* the *professional development* of therapists - as time passes, difficult patients stay difficult, difficult circumstances do not get any easier, but difficulties located in therapists themselves abate. The studies reported in this thesis investigated the dimensions and correlates of the *pervasiveness* of experiences of difficulties, both within therapists (*specificity*) and between therapists (*universality*); initially conceptualised as a spectrum ranging from transient to paradigmatic, and thought to be associated with familiarity, impact, and duration of difficulties.

In retrospect, this heuristic appears somewhat naive, however, it provided a starting point. The subsequently developed definitions of transient, paradigmatic, and situational difficulty types in effect collapsed two of the previously established locations into one - difficult patients and difficult circumstances become components of situational difficulties; while differentiating the other - difficulties located within the therapist are divided into those arising from competency deficits and those due to idiosyncratic personal characteristics. These concepts have been validated in two subsequent studies.

As far as situational difficulties are concerned, we now know a little more about what characteristics lead patients to be perceived as difficult - their high levels of

interpersonal problems, their general lack of co-operation and their controlling hostility. This confirms by and large what previous studies had found. We also have some information about situational difficulties in general - therapists see them as universal; view the patients with whom they are occurring as dissimilar to themselves and as attacking; but perceive themselves as staying in role during such difficulties and meeting the challenges in an affiliative and autonomy promoting manner, while managing to stay on good terms with themselves. They are tested but not troubled.

About transient difficulties we have learned that they are remarkably unrelated to therapists' personal characteristics; that they are more likely to occur for not-so-experienced therapists; are felt as unfamiliar when encountered; and perceived as related to deficits in therapists' skills, knowledge and experience. They seem to fit well into the model of professional development amplified by Skovholt & Rønnestad (1995), and their resolution can be expected to contribute to such development.

Much has also been learned about paradigmatic difficulties. They have a high emotional impact; are viewed as related to problems occurring outside the therapeutic setting, and thus characteristic of the therapist experiencing them; are more likely to occur with patients whom therapists perceive as similar to themselves; arouse a wish for hostile control in the therapist in the absence of perceived patient hostility; evoke internal states in therapists which are similar to their worst self-representations and dissimilar to their best; and are related to therapists internalised attachment conflicts. They appear to call for a model of personal development occurring alongside and interwoven with professional development. To develop such a model is a challenge for the future.

What of the consequences of these findings? Different types of difficulties call for differential responses both in internal reflection and in the supervisory setting. Situational difficulties require tolerance and acceptance; transient difficulties call for an improved repertoire of knowledge and skills, or opportunities for wider experience; and paradigmatic difficulties need enhanced self-awareness. As difficulties are likely to be a mixture of two or all three types, supervisory skill may well be related to the ability to give differential responses to the respective components and refer them to the appropriate resources - for instance, theoretical formulation, private reflection, personal therapy, reading, support, or discussions with peers. Questions helping in disentangling the various components can be derived from difficulty type definitions, for instance, 'have you come across such a difficulty outside the therapeutic setting? - with other patients? - in the past?', or 'would another therapist find this situation difficult?'

The findings also have potential consequences for clinical theory. The challenge there is to bridge the gap between low-level formulations which stay close to the empirical data but may be simplistic, and concepts which do justice to clinical complexities but require high levels of theoretical inference. To do so would be outside the scope of this thesis, which aimed to stay in its theoretical formulations near to the observations derived from its three studies.

The wider consequences, however, concern development. My own speculative view in this respect is that paradigmatic difficulties are best understood as attempts at mastery, as striving to work on internal conflict in the service of personal development. Therapists enter the profession for many different reasons, though the wish to rearrange and improve internal representations of the self and others is, I believe,

prominent among them. Patients, as Searles (1975) contends, may well play a complementary role in this pursuit. From such a viewpoint, absence of paradigmatic difficulties would be indicative of personal stagnation, premature closure, or selective avoidance. Their presence would indicate developmental opportunities; the way in which they are resolved decide, how well such opportunities are taken. In the end, professional and personal development have to go hand in hand.

LIST OF REFERENCES:

(prepared in accordance with Psychology Department conventions)

Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment, 55*, 521-536.

Alpher, V. S. (1991). Interdependence and parallel processes: A case study of structural analysis of social behavior in supervision and short-term dynamic psychotherapy. *Psychotherapy, 28*, 218-231.

Alpher, V. S. (1996). Identity and introject in dissociative disorders. *Journal of Consulting and Clinical Psychology, 64*, 1238-1244.

Bandura, A., Lipsher, D. H., & Miller, P. E. (1960). Psychotherapists' approach-avoidance reactions to patients' expressions of hostility. *Journal of Consulting Psychology, 24*, 1-7.

Barber, J. P., Crits-Cristoph, P., & Luborsky, L. (1990). A guide to the CCRT standard categories and their classification. In L. Luborsky & P. Crits-Christoph (Eds.), *Understanding transference. The CCRT method* . New York: Basic Books.

Barker, C., Pistrang, N., & Elliott, R. (1994). *Research methods in clinical and counselling psychology*. Chichester: Wiley.

Barkham, M., Hardy, G. E., & Startup, M. (1994). The structure, validity and clinical relevance of the Inventory of Interpersonal Problems. *British Journal of Medical Psychology, 67*, 171-185.

Barkham, M., Hardy, G. E., & Startup, M. (1996). The IIP-32: A short version of the Inventory of Interpersonal Problems. *British Journal of Clinical Psychology, 35*, 21-35.

Barkham, M., Mellor-Clark, J., Evans, C., Margison, F., McGrath, G., Audin, K., & Connell, J. (1998). Efficacy and effectiveness in the context of psychotherapy research generations. In CORE System Group (Ed.), *CORE system (information management) handbook* . Leeds: CORE System Group.

Basch, M. F. (1980). *Doing psychotherapy*. New York: Basic Books.

Beck, D. F. (1988). *Counselor characteristics: How they affect outcomes*. Milwaukee: Family Service America.

Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review, 81*, 392-425.

Benjamin, L. S. (1979). Use of Structural Analysis of Social Behavior (SASB) and Markov chains to study dyadic interactions. *Journal of Abnormal Psychology, 88*, 303-319.

Benjamin, L. S. (1982). Use of Structural Analysis of Social Behavior (SASB) to guide interventions in psychotherapy. In J. Anchin & D. Kiesler (Eds.), *Handbook of interpersonal psychotherapy* . New York: Pergamon Press.

Benjamin, L. S. (1983). The INTREX user's manual. Madison: INTREX Interpersonal Institute, Inc.

Benjamin, L. S. (1984). Principles of prediction using Structural Analysis of Social Behavior. In R. A. Zucker, J. Aronoff, & A. J. Rabin (Eds.), *Personality and the prediction of behavior.* New York: Academic Press.

Benjamin, L. S. (1986). Operational definition and measurement of dynamics shown in the stream of free association. *Psychiatry, 49*, 104-129.

Benjamin, L. S. (1987). Use of the SASB dimensional model to develop treatment plans for personality disorders. I: Narcissism. *Journal of Personality Disorders, 1*, 43-70.

Benjamin, L. S. (1988a). SASB INTREX short form user's manual . Salt Lake City: INTREX Interpersonal Institute, Inc.

Benjamin, L. S. (1988b). SASB short form, MS-DOS computer software. Madison: INTREX Interpersonal Institute, Inc.

Benjamin, L. S. (1993). Every psychopathology is a gift of love. *Psychotherapy Research, 3*, 1-24.

Benjamin, L. S. (1994a). Good defenses make good neighbors. In H. R. Conte & R. Plutchik (Eds.), *Ego defenses: Theory and measurement*. New York: Wiley.

Benjamin, L. S. (1994b). SASB: A bridge between personality theory and clinical psychology. *Psychological Inquiry, 5*, 273-316.

Benjamin, L. S. (1995). SASB Intrex Short Form user's manual. Salt Lake City: University of Utah.

Benjamin, L. S. (1996a). A clinician-friendly version of the Interpersonal Circumplex: Structural Analysis of Social Behavior (SASB). *Journal of Personality Assessment, 66*, 248 - 266.

Benjamin, L. S. (1996b). Introduction to the special section on Structural Analysis of Social Behavior. *Journal of Consulting and Clinical Psychology, 64*, 1203-1212.

Benjamin, L. S. (1997). SASB short form, Windows 95 computer software. Salt Lake City: University of Utah.

Benjamin, L. S., Foster, S. W., Giat-Roberto, L., & Estroff, S. E. (1986). Breaking the family code: Analysis of videotapes of family interactions by Structural Analysis of

Social Behavior (SASB). In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* . London: Guilford.

Benjamin, L. S. & Friedrich, F. J. (1991). Contributions of Structural Analysis of Behavior (SASB) to the bridge between cognitive science and a science of object relations. In M. J. Horowitz (Ed.), *Person schemas and maladaptive interpersonal patterns*. Chicago: The University of Chicago Press.

Berzofsky, M., Freedman, N., Essig, T., & Margolis, R. (1992). *The shape of the communicated transference in difficult and not so difficult patients. Paper III in: Difficult moments in difficult sessions with difficult patients.* Paper presented at the Society for Psychotherapy Research, 23rd Annual Meeting, Berkley.

Beutler, L. E., Crago, M., & Arizmendi, T. G. (1986). Therapist variables in psychotherapy process and outcome. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd edition). New York: Wiley.

Beutler, L. E., Machado, P. P., & Neufeldt, S. A. (1994). Therapist Variables. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th edition). New York: Wiley.

Binns, M., Davis, J. D., Davis, M. L., Elliott, R., Francis, V. M., Kelman, J. E., & Schröder, T. A. (1986). *Taxonomy of therapist difficulties: Rating manual.* Coventry: University of Warwick.

Birtchnell, J. (1994). The interpersonal octagon: An alternative to the interpersonal circle. *Human Relations, 47*, 511-527.

Blos, P. (1962). *On Adolescence*. New York: Free Press.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy, 16*, 252-260.

Borkenau, P. & Amelang, M. (1985). The control of social desirability in personality inventories. A study using the principal-factor deletion technique. *Journal of Research in Personality, 19*, 44-53.

Botermans, J.-F., Orlinsky, D. E., & Wiseman, H. (1998). *The therapist's own therapy in international perspective: Impact on personal wellbeing and professional development.* Paper presented at the Society for Psychotherapy Research, 29th Annual Meeting, Snowbird (UT).

Browning, D. C. (1972). *Everyman's thesaurus of English words and phrases*. London: Pan Books.

Carson, R. C. (1969). *Interaction concepts of personality*. Chicago: Aldine.

Cecchin, G. (1987). Hypothesizing, circularity and neutrality revisited: An invitation to curiosity. *Family Process, 26*, 405-413.

Clementel Jones, C., Azzone, P., Battacchi, M. W., & Freni, S. (1996). Italian translation of the Inventory of Interpersonal Problems, validity and stability of factor structure. *New Trends in Experimental and Clinical Psychiatry, 12*, 253-259.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304-1312.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Coltart, N. E. C. (1986). 'Slouching towards Bethlehem...' or thinking the unthinkable in psychoanalysis. In G. Kohon (Ed.), *The British school of psychoanalysis. The independent tradition*. London: Free Association Books.

Crits-Christoph, P., Barber, J. P., Baranackie, K., & Cooper, A. (1993). Assessing the therapist's interpretations. In N. E. Miller, L. Luborsky, J. P. Barber, & J. P. Docherty (Eds.), *Psychodynamic treatment research*. New York: Basic Books.

Crits-Cristoph, P. & Demorest, A. (1988). List of standard categories (edition 2). Pittsburgh: University of Pennsylvania.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Crook, J. H. (1980). *The evolution of human consciousness*. Oxford: Oxford University Press.

Cutler, R. L. (1958). Countertransference effects in psychotherapy. *Journal of Consulting Psychology, 22*, 349-356.

Davies-Osterkamp, S., Hartkamp, N., & Junkert, B. (1993). Die Intrex-Kurzform. In W. Tress (Ed.), *Die strukturale Analyse sozialen Verhaltens*. Heidelberg: Asanger.

Davis, J. D., Davis, M. L., Davidson, C., & Schröder, T. A. (1990). *The development of psychotherapists.* Paper presented at the Society for Psychotherapy Research (UK), 7th Annual Meeting, Ravenscar.

Davis, J. D., Davis, M. L., Schröder, T. A., Orlinsky, D., & Howard, K. (1992). *Therapists' experiences of difficulties in practice and their relation to the psychotherapist's level of experience.* Paper presented at the Society for Psychotherapy Research, 23rd Annual Meeting, Berkeley.

Davis, J. D., Elliott, R., Davis, M. L., Binns, M., Francis, V. M., Kelman, J. E., & Schröder, T. A. (1987). Development of a taxonomy of therapists' difficulties: Initial report. *British Journal of Medical Psychology, 60*, 109-119.

Davis, J. D., Francis, V. M., Davis, M. L., Schröder, T. A., Kelman, J. E., Binns, M., & Elliott, R. (1989). *Development of a taxonomy of therapists' coping strategies: Initial report* (Unpublished manuscript). Coventry: University of Warwick.

Davis, J. D., Schröder, T. A., Davis, M. L., & Orlinsky, D. E. (1996). *Therapists' difficulties in practice: Dimensions and relationships to therapists' experience levels.* Paper presented at the Society for Psychotherapy Research, 27th Annual Meeting, Amelia Island (FL).

Derogatis, L. R. (1975). *Brief Symptom Inventory*. Baltimore: Clinical Psychometric Research.

Deutsch, C. J. (1984). Self-reported sources of stress among psychotherapists. *Professional Psychology: Research and Practice, 15*, 833-845.

Deyhle, D. L., Hess, G. A. J., & LeCompte, M. D. (1992). Approaching ethical issues for qualitative researchers in education. In M. D. LeCompte, W. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education*. New York: Academic Press.

Elkind, S. N. (1992). *Resolving impasses in therapeutic relationships*. New York: Guilford.

Elliott, R. (1989). *Issues in the selection, management and training of raters.* Paper presented at the Society for Psychotherapy Research, 20th Annual Meeting, Toronto.

Erikson, E. H. (1980). Elements of a Psychoanalytic Theory of Psychosocial Development. In S. I. Greenspan & G. H. Pollock (Eds.), *The Course of Life: Psychoanalytic Contributions Toward Understanding Personality Development (Vol. 1: Infancy and Early Childhood).* Rockville: National Institute of Mental Health.

Essex, M., Klein, M., Lohr, M., & Benjamin, L. S. (1985). Intimacy and depression in older women. *Psychiatry, 48*, 159-178.

Everitt, B. S. (1996). *Making sense of statistics in psychology*. Oxford: Oxford University Press.

Farber, B. A. (1983a). The effects of psychotherapeutic practice on psychotherapists. *Psychotherapy: Theory, Research, and Practice, 20*, 174-182.

Farber, B. A. (1983b). Psychotherapists' perceptions of stressful patient behaviour. *Professional Psychology: Research and Practice, 14*, 697-705.

Farber, B. A. & Heifetz, L. J. (1981). The satisfactions and stresses of psychotherapeutic work: A factor analytic study. *Professional Psychology: Research and Practice, 12*, 621-630.

Fiedler, F. E. (1951). A method of objective quantification of certain countertransference attitudes. *Journal of Clinical Psychology, 7*, 101-107.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-358.

Fonagy, P. & Moran, G. (1993). Selecting single case research designs for clinicians. In N. E. Miller, L. Luborsky, J. P. Barber, & J. P. Docherty (Eds.), *Psychodynamic treatment research*. New York: Basic Books.

France, A. C. I. & Alpher, V. S. (1995). Structural analysis of social behavior and perceptions of caregiving. *Journal of Psychology, 129*, 375-388.

Freedman, N. (1992a). *Difficult moments in difficult sessions with difficult patients.* Paper presented at the Society for Psychotherapy Research, 23rd Annual Meeting, Berkeley.

Freedman, N. (1992b). *Introduction and approach. Paper I and II in: Difficult moments in difficult sessions with difficult patients.* Paper presented at the Society for Psychotherapy Research, 23rd Annual Meeting, Berkeley.

Freedman, N., Lavender, J., & Berger, M. (1992). *On receiving the patient's transference: Symbolized and desymbolized countertransference. Paper IV in: Difficult moments in difficult sessions with difficult patients.* Paper presented at the Society for Psychotherapy Research, 23rd Annual Meeting, Berkeley.

Freedman, N., Lavender, J., & Hoffennberg, J. (1992). *On resymbolization: The process of disembedding. Paper V in: Difficult moments in difficult sessions with difficult patients.* Paper presented at the Society for Psychotherapy Research, 23rd Annual Meeting, Berkeley.

Fremont, S. K. & Anderson, W. (1988). Investigation of factors involved in therapists' annoyance with clients. *Professional Psychology: Research and Practice, 19*, 330-335.

Gilbert, P. (1989). *Human nature and suffering*. London: Erlbaum.

Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Gower, J. C. (1985). Measures of similarity, dissimilarity, and distance. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 5). New York: Wiley.

Greenspan, M. & Kulish, N. M. (1985). Factors in premature termination in long-term psychotherapy. *Psychotherapy, 22*, 75-82.

Gurtman, M. B. (1992). Construct validity of interpersonal personality measures: The interpersonal circumplex as a nomological net. *Journal of Personality and Social Psychology, 63*, 105-118.

Gurtman, M. B. (1997). Studying personality traits: The circular way. In R. Plutchik & H. R. Conte (Eds.), *Circumplex models of personality and emotions*. Washington: American Psychological Association.

Hartley, D. (1991). Assessing interpersonal behavior patterns using Structural Analysis of Social Behavior (SASB). In M. J. Horowitz (Ed.), *Person schemas and maladaptive interpersonal patterns*. Chicago: The University of Chicago Press.

Henry, W. P. (1996). Structural analysis of social behavior as a common metric for programmatic psychopathology and psychotherapy research. *Journal of Consulting and Clinical Psychology, 64*, 1263-1275.

Henry, W. P., Schacht, T. E., & Strupp, H. (1986). Structural Analysis of Social Behavior: Application to a study of interpersonal process in differential psychotherapeutic outcome. *Journal of Consulting and Clinical Psychology, 54*, 27-31.

Henry, W. P., Schacht, T. E., & Strupp, H. (1990). Patient and therapist introject, interpersonal process, and differential psychotherapy outcome. *Journal of Consulting and Clinical Psychology, 58*, 768-774.

Hicks, C. & Nixon, S. (1989). The use of a modified Repertory Grid technique for assessing the self-concept of children in Local Authority foster care. *British Journal of Social Work, 19*, 203-216.

Hogan, R. A. (1964). Issues and Approaches in Supervision. *Psychotherapy, 1*, 139-141.

Horowitz, L., French, R., Gani, M., & Lapid, J. (1980). The co-occurence of semantically similar interpersonal problems. *Journal of Consulting and Clinical Psychology, 48*, 413-415.

Horowitz, L., Weckler, D., & Doren, R. (1983). Interpersonal problems and symptoms: A cognitive approach. In P. Kendall (Ed.), *Advances in cognitive-behavioral research and therapy*. New York: Guilford.

Horowitz, L. M. (1979). On the cognitive structure of interpersonal problems treated in psychotherapy. *Journal of Consulting and Clinical Psychology, 47*, 5-15.

Horowitz, L. M. (1996). The study of interpersonal problems: A Leary legacy. *Journal of Personality Assessment, 66*, 283-300.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1987). The Inventory of Interpersonal Problems. Unpublished manuscript. Stanford: Stanford University.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885-892.

Horowitz, M. J., Wilner, M., & Alvarez, W. (1979). Impact of event scale: A measure of subjective stress. *Psychosomatic Medicine, 41*, 209-218.

Hunt, J. C. (1989). *Psychoanalytic aspects of fieldwork*. London: Sage.

Ichiyama, M. A., Zucker, R. A., Fitzgerald, H. E., & Bingham, C. R. (1996). Articulating subtype differences in self and relational experience among alcoholic men using Structural Analysis of Social Behavior. *Journal of Consulting and Clinical Psychology, 64*, 1245-1254.

Jackson, D. & Helmes, E. (1979). Personality structure and the circumplex. *Journal of Personality and Social Psychology, 37*, 2278-2285.

Kächele, H. & Dahlbender, R. W. (1993). Übertragung und zentrale Beziehungsmuster. In P. Buchheim, M. Cierpka, & T. Seifert (Eds.), *Beziehung im Fokus, Weiterbildungsforschung*. Heidelberg: Springer-Verlag.

Kellerman, J. (1985). *When the bough breaks*. London: Warner Books.

Kelly, G. A. (1955). *The psychology of personal constructs*. New York: Norton.

Kernberg, O. F. (1984). *Severe personality disorders. Psychotherapeutic strategies*. New Haven: Yale University Press

Kiesler, D. J. (1973). *The process of psychotherapy*. Chicago: Aldine.

Kiesler, D. J., Anchin, J. C., Perkins, M. J., Chirico, B. M., Kyle, E. M., & Federman, E. J. (1976). The impact message inventory. Richmond: Virginia Commonwealth University.

Kiesler, D. J., Schmidt, J. A., & Wagner, C. (1997). A circumplex inventory of impact messages: An operational bridge between emotion and interpersonal behavior. In R. Plutchik & H. R. Conte (Eds.), *Circumplex models of personality and emotions*. Washington: American Psychological Association.

Korchin, S. J. & Cowan, P. (1982). Ethical perspectives in clinical research. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* . New York: Wiley.

Krzanowski, W. J. (1988). *Principles of multivariate analysis*. Oxford: Clarendon Press.

Lambert, M. J. & Hill, C. (1994). Assessing psychotherapy outcomes and process. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th edition). New York: Wiley.

Langs, R. (1973). *The technique of psychoanalytic psychotherapy*. (Volume 1). London: Jason Aronson.

Leary, T. (1957). *Interpersonal diagnosis of personality*. New York: Ronald.

Loganbill, C., Hardy, E., & Delworth, U. (1982). Supervision: A conceptual model. *Counseling Psychologist, 10*, 3-42.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lovie, P. (1991). Regression diagnostics. A rough guide to safer regression. In P. Lovie & A. D. Lovie (Eds.), *New developments in statistics for psychology and the social sciences* (Vol. 2). London: BPS / Routledge.

Luborsky, L. (1976). Helping alliances in psychotherapy: The groundwork for a study of their relationship to its outcome. In J. L. Claghorn (Ed.), *Successful psychotherapy* . New York: Brunner/Mazel.

Luborsky, L. (1977). Measuring a pervasive psychic structure in psychotherapy: The core conflictual relationship theme. In N. Freedman & S. Grand (Eds.), *Communicative structures and psychic structures*. New York: Plenum Press.

Luborsky, L. (1986). *A set of standard categories for the CCRT. Edition 1*. Unpublished manuscript.

Luborsky, L. (1990a). The early development of the core conflictual relationship theme idea. In L. Luborsky & P. Crits-Christoph (Eds.), *Understanding transference. The CCRT method*. New York: Basic Books.

Luborsky, L. (1990b). A guide to the CCRT method. In L. Luborsky & P. Crits-Christoph (Eds.), *Understanding transference. The CCRT method* . New York: Basic Books.

Luborsky, L., Crits-Christoph, P., & Mellon, J. (1986). Advent of Objective Measures of the Transference Concept. *Journal of Consulting and Clinical Psychology, 54*, 39-47.

Luborsky, L., Crits-Christoph, P., Mintz, J., & Auerbach, A. (1988). *Who will benefit from psychotherapy? Predicting therapeutic outcomes*. New York: Basic Books.

Luborsky, L. & Crits-Cristoph, P. (1989). A relationship pattern measure: The core conflictual relationship theme. *Psychiatry, 52*, 250-259.

Luborsky, L. & Kächele, H. (1988). *Der zentrale Beziehungskonflikt. Ein Arbeitsbuch*. Ulm: PSZ-Verlag.

Luborsky, L., Mellon, J., Cohen, K. D., Ravenswaay, P., Hole, A. V., Childress, A. R., Ming, S., Crits-Cristoph, P., Levine, F. J., & Alexander, K. (1985). A verification of Freud's grandest clinical hypothesis: The transference. *Clinical Psychology Review, 5*, 231-246.

Luborsky, L., Popp, C., & Barber, J. P. (1994). Common and special factors in different transference-related measures. *Psychotherapy Research, 4*, 277-286.

Malcolm, J. (1988). *Psychoanalysis: The impossible profession*. London: Maresfield.

Maslach, C. (1982). *The cost of caring*. Englewood Cliffs: Prentice Hall

Mathers, N., Jones, N., & Hannay, D. (1995). Heartsink patients: A study of their general practitioners. *British Journal of General Practice, 45*, 293-296.

Mattinson, J. (1975). *The reflection process in casework supervision*. London: Institute of Marital Studies.

McCrae, R. R. & Costa, P. T. J. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 822-838.

Meares, R. A. & Hobson, R. F. (1977). The persecutory therapist. *British Journal of Medical Psychology, 50*, 349-359.

Medeiros, M. E. & Prochaska, J. O. (1988). Coping strategies that therapists use in working with stressful clients. *Professional Psychology: Research and Practice, 19*, 112-114.

Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis*. (2nd edition). London: Sage.

Milne, D. (1990). Assessment: A mental health portfolio . Windsor: NFER-Nelson.

Mittwoch, A. (1990). Commentary on 'Crowding or the complexity of group relationships'. *Group Analysis, 23*, 325-327.

Modell, A. H. (1990). *Other times, other realities. Towards a theory of psychoanalytic treatment*. Cambridge (MA): Harvard University Press.

Moras, K. & Hill, C. E. (1991). Rater selection for psychotherapy process research: An evaluation of the state of the art. *Psychotherapy Research, 1*, 113-123.

Neimeyer, R. A., Klein, M. H., Gurman, A. S., & Greist, J. H. (1983). Cognitive structure and depressive symptomatology. *British Journal of Cognitive Psychotherapy, 1*, 65-73.

Norcross, J. C. & Prochaska, J. O. (1986). Psychotherapist heal thyself. *Psychotherapy, 23*, 345-356.

Norcross, J. C. & Wogan, M. (1983). American psychotherapists of diverse persuasions: Characteristics, theories, practices, and clients. *Professional Psychology: Research and Practice, 14*, 529-539.

Nunnally, J. C. (1978). *Psychometric theory*. (2nd Edition). New York: McGraw-Hill.

Orlinsky, D. E. & Howard, K. I. (1977). The therapist's experience of psychotherapy. In A. S. Gurman & A. M. Razin (Eds.), *Effective psychotherapy: A handbook of research*. Oxford: Pergamon Press.

Orlinsky, D. E. & Howard, K. L. (1975). *Varieties of psychotherapeutic experience*. New York: Teachers College Press.

Palazzoli-Selvini, M., Boscolo, L., Cecchin, G., & Prata, G. (1980). Hypothesizing - Circularity - Neutrality: Three guidelines for the conductor of the session. *Family Process, 19*, 3-12.

Pearlman, L. A. & Saakvitne, K. W. (1995). *Trauma and the therapist*. London: Norton.

Plutchik, R., Conte, H. R., & Karasu, T. B. (1994). Critical incidents in psychotherapy. *American Journal of Psychotherapy, 48*, 75-84.

Pope, K. S. & Tabachnik, B. G. (1994). Therapists as patients: A national survey of psychologists' experiences, problems, and beliefs. *Professional Psychology: Research and Practice, 25*, 247-258.

Quintana, S. M. & Meara, N. M. (1990). Internalization of therapeutic relationships in short-term psychotherapy. *Journal of Counseling Psychology, 37*, 123-130.

Real, T. (1990). The therapeutic use of self in constructionist / systemic therapy. *Family Process, 29*, 255-272.

Rioch, M. J., Coulter, W. R., & Weinberger, D. M. (1976). *Dialogues for therapists*. San Francisco: Jossey-Bass.

Robbins, J. M., Beck, P. R., Mueller, D. P., & Mizener, D. A. (1988). Therapists' perceptions of difficult psychiatric patients. *Journal of Nervous and Mental Disease, 176*, 490-497.

Rodolfa, E. R., Kraft, W. R., & Reilly, R. R. (1988). Stressors of professionals and trainees at APA-approved counseling and VA medical center internship sites. *Professional Psychology: Research and Practice, 19*, 43-49.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 45*, 775-777.

Russell, R. L. & Stiles, W. B. (1979). Categories for classifying language in psychotherapy. *Psychological Bulletin, 86*, 404-419.

Rycroft, C. (1972). *A critical dictionary of psychoanalysis*. Harmondsworth: Penguin.

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric qualities of rating data. *Psychological Bulletin, 88*, 413-428.

Sandler, J., Dare, C., & Holder, A. (1973). *The patient and the analyst: The basis of the psychoanalytic process*. London: Allen and Unwin.

Schaefer, E. S. (1965). A configurational analysis of children's reports of parent behavior. *Journal of Consulting Psychology, 29*, 552-557.

Schauenburg, H., Pekrun, G., & Leibing, E. (1995). Diagnostik interpersonaler Probleme bei depressiven Störungen. (The assessment of interpersonal problems in depressive disorders.). *Zeitschrift für Klinische Psychologie, Psychopathologie und Psychotherapie, 43*, 200-213.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350-353.

Schröder, T. A. (1993). *Varieties of therapists' experiences of difficulties in practice.* Paper presented at the Society for Psychotherapy Research, 4th European Meeting, Budapest.

Schröder, T. A. (1996). *Was lange währt wird auch nicht leichter. Schwierigkeiten und Bewältigungsstrategien in der psychotherapeutischen Arbeit.* Paper presented at the Lindauer Psychotherapie Wochen, Lindau.

Schröder, T. A. (1997). Was lange währt wird auch nicht leichter - Schwierigkeiten und Bewältigungsstrategien in der psychotherapeutischen Arbeit. In P. L. Janssen, M. Cierpka, & P. Buchheim (Eds.), *Psychotherapie als Beruf.* Goettingen: Vandenhoeck & Ruprecht.

Schröder, T. A., Binns, M., Davis, J. D., Davis, M. L., Elliott, R., Francis, V. M., & Kelman, J. E. (1987a). A taxonomy of therapist coping strategies. In H. Kaechele, M. Cierpka, & M. Hölzer (Eds.), *Society for Psychotherapy Research, Abstracts of the 18th Annual Meeting*. Ulm: PSZ-Verlag.

Schröder, T. A., Binns, M., Davis, J. D., Davis, M. L., Elliott, R., Francis, V. M., & Kelman, J. E. (1987b). *A taxonomy of therapist difficulties.* In H. Kaechele, M. Cierpka, & M. Hölzer (Eds.), *Society for Psychotherapy Research, Abstracts of the 18th Annual Meeting*. Ulm: PSZ-Verlag.

Schröder, T. A. & Davis, J. D. (1993). *The new and the short and the tall. Varieties of therapists' experiences of difficulties.* Paper presented at the Society for Psychotherapy Research (UK), 10th Annual Meeting, Ravenscar.

Schröder, T. A. & Davis, J. D. (1994). *Dimensions of difficulties. How can psychotherapists' experiences of problems in their practice be stratified?* Paper presented at the Society for Psychotherapy Research, 25th Annual Meeting, York.

Schröder, T. A., Willutzki, U., & Ambuehl, H. R. (1994). *Schwierigkeiten und Bewältigungsstrategien in der psychotherapeutischen Arbeit.* Paper presented at the Deutsche Gesellschaft für Psychologie, 39. Jahreskongress, Hamburg.

Searles, H. F. (1975). The patient as therapist to his analyst. In P. L. Giovacchini (Ed.), *Tactics and techniques in psychoanalytic therapy* (Vol. 2). New York: Jason Aronson.

Searles, H. F. (1979). *Countertransference and related subjects*. New York: International Universities Press.

Shlien, J. M. (1987). A countertheory of transference. *Person-Centred Review, 2*, 15-49.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428

Sieber, J. E. (1992). *Planning ethically responsible research: A guide for students and internal review boards*. Newbury Park: Sage.

Singer, B. A. & Luborsky, L. (1977). Countertransference: The status of clinical versus quantitative research. In A. S. Gurman & A. M. Razin (Eds.), *Effective psychotherapy: A handbook of research*. Oxford: Pergamon Press.

Singer, J. L., Sincoff, J. B., & Kolligian, J. (1989). Countertransference and cognition: Studying the psychotherapist's distortions as consequences of normal information processing. *Psychotherapy, 26*, 344-355.

Skovholt, T. M. & Rønnestad, M. H. (1992). Themes in Therapist and Counselor Development. *Journal of Counseling & Development, 70*, 505-515.

Skovholt, T. M. & Rønnestad, M. H. (1995). *The evolving professional self: Stages and themes in therapist and counselor development*. Chichester: Wiley.

Slater, P. (1976). *The measurement of intrapersonal space by grid technique*. London: Wiley.

Snyder, W. U. (1961). *The psychotherapy relationship*. New York: Macmillan.

Soldz, S., Budman, S., Demby, A., & Merry, J. (1995). A short form of the Inventory of Interpersonal Problems circumplex scales. *Psychological Assessment, 2*, 53-63.

Stoltenberg, C. & Delworth, U. (1987). *Supervising counselors and therapists: A developmental approach*. San Francisco: Jossey-Bass.

Strauss, B. M. (1994). IIP-C Deutsche Ausgabe. Weinheim: Beltz Test GmbH.

Strauss, B. M. & Kaechele, H. (1998). The writing on the wall - comments on the current discussion about empirically validated treatments in Germany. *Psychotherapy Research, 8*, 158-170.

Strupp, H. H. (1995). *The therapist as a key variable in dynamic psychotherapy.* Paper presented at the Society for Psychotherapy Research, 26th Annual Meeting, Vancouver.

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. New York: Norton.

Sussman, M. B. (1995). *A perilous calling. The hazards of psychotherapeutic practice.* New York: Wiley.

Symington, N. (1983). The analyst's act of freedom as an agent of therapeutic change. *International Review of Psycho-Analysis, 10*, 283-291.

Tabachnik, B. G. & Fidell, L. S. (1989). *Using multivariate statistics*. (2nd edition). New York: Harper & Row.

Torres, M. A. (1983). *Conscious countertransference reactions by beginning therapists in response to patient anger.* Unpublished PhD Dissertation. New York: Adelphi University.

Tracey, T. J. G., Rounds, J., & Gurtman, M. (1996). Examination of the general factor with the interpersonal circumplex structure: Application to the Inventory of Interpersonal Problems. *Multivariate Behavioral Research, 31*, 441-466.

Tress, W. (1993). *Die Strukturale Analyse Sozialen Verhaltens - SASB*. Heidelberg: Asanger.

Tress, W. & Junkert, B. (1993). Das SASB Modell - Modellstruktur. In W. Tress (Ed.), *Die Strukturale Analyse Sozialen Verhaltens - SASB*. Heidelberg: Asanger.

Tscheulin, D. & Glossner, A. (1993). Der Intrex-Fragebogen. In W. Tress (Ed.), *Die strukturale Analyse sozialen Verhaltens - SASB.* Heidelberg: Asanger.

Vaillant, G. E. (1976). Natural History of Male Psychological Health: The Relation of Choice of Ego Mechanisms of Defense to Adult Adjustment. *Archives of General Psychiatry, 33*, 535-545.

Weinryb, R. M., Gustavsson, J. P., Hellstrom, C., & Anderson, E. (1996). Interpersonal problems and personality characteristics: Psychometric studies of the Swedish version of the IIP. *Personality and Individual Differences, 20*, 13-23.

Weiss, D. J., Horowitz, M. J., & Wilner, N. (1984). The stress response rating scale: A clinician's measure for rating the response to serious life-events. *British Journal of Clinical Psychology, 23*, 202-215.

Wiggins, J. S., Steiger, J. H., & Gaelick, L. (1981). Evaluating circumplexity in personality data. *Multivariate Behavioral Research, 16*, 263-289.

Wiggins, J. S., Trapnell, P., & Phillips, N. (1988). Psychometric and geometric characteristics of the revised Interpersonal Adjective Scale (IAS-R). *Multivariate Behavioral Research, 23*, 517-530.

Wilkinson, L. (1996). *SYSTAT 6.0 for Windows: Statistics Manual*. Chicago: SPSS INC.

Winer, B. J. (1971). *Statistical principles in experimental design*. (2nd edition). New York: McGraw-Hill.

Winter, D. A. (1992). *Personal construct psychology in clinical practice*. London: Routledge.

Woolfe, R. (1996). The nature of counselling psychology. In R. Woolfe & W. Dryden (Eds.), *Handbook of counselling psychology* . London: Sage.

KEY TO ABBREVIATIONS:

ANOVA	Analysis of Variance
ASCII	American Standard Code for Information Interchange
BPS	British Psychological Society
CCQ	Common Core Questionnaire
BSI	Brief Symptom Inventory
CCRT	Core Conflictual Relationship Theme
CPTC	Counselling in Primary Care Trust
CRN	Collaborative Research Network
DOS	Disk Operating System
DSM	Diagnostic and Statistical Manual of mental disorders
E	East
FID	Familiarity/Impact/Duration
GP	General Practitioner
IIP	Inventory of Interpersonal Problems
IIP-C	Inventory of Interpersonal Problems-Circumplex
IIP-SC	Inventory of Interpersonal Problems- Short Circumplex
IMI	Impact Message Inventory
IPC	Interpersonal Circle
ISPDP	International Study of the Professional Development of Psychotherapists
LPW	Lindauer Psychotherapie Wochen
MSc	Master of Science
N	North
NAPP	Nurses Association for Psychodynamic Psychotherapy
NE	North East

NHS	National Health Service
NW	North West
OCD	Obsessive-compulsive Disorder
PTSD	Post-traumatic Stress Disorder
R	Rater
RTS	Returned to sender
S	South
SASB	Structural Analysis of Social Behaviour
SDMHT	Southern Derbyshire Mental Health (NHS) Trust
SE	South East
SPR	Society for Psychotherapy Research
SW	South West
TDQ	Therapist Difficulty Questionnaire
UK	United Kingdom
UKCP	United Kingdom Council for Psychotherapy
W	West